

-2099-

++G +S G+IL+HG+V+ E T FNGI I GA + +Q RVLMLS+KAR DANFIL
 Sbjct: 306 HGGKHSEGYILKHGVNREAAATSI FNGISKIEHGATKSHGQTERVLMISEKARGDANFIL 365
 Query: 349 LIDENDVTAGHAASIGQVDPEDLYYJMSRGLNQKTAELQVIRGFLGTIVIAEIPVKEVRDE 408
 LIDE+DVTAGHAAS+G+DP ++YLMERG++ AE+LVI GFL V+ ++P++ V++
 Sbjct: 366 LIDEDDVTAGHAASVGKIDPIQMPLYMSRGISRAEAERLVINGFLAPVVGQLPVSVKER 425
 Query: 409 MIAVIDTKLE 418
 ++ I+ K++
 Sbjct: 426 LVEAIERKVK 435

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5783> which encodes the amino acid sequence <SEQ ID 5784>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.80 Transmembrane 387 - 403 (387 - 403)

----- Final Results -----

bacterial membrane --- Certainty=0.1319 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CB15259 GB:299120 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 177/428 (41%), Positives = 267/428 (62%), Gaps = 15/428 (3%)
 Query: 3 KRKLVPASQAHAPAWLQERRLAALAEIPNLEPTIERVKFHRWNLGDGT--LTENESLA 60
 +E L +PS+ H BPWL+ RL ALE +L +P ++ K WN + +NE L+
 Sbjct: 11 QETLSPFSEKHQEPAWLNLRQLALEQAEADLMPKPDKTKITNWNFINFAKETVDNEPL 70
 Query: 61 SVEDF-----IAIGDNPKLVQVGQTQTVLEQLPMA--LIDKGVFSDPTALEEIPVI 111
 S+D I I + K + V L ++ L DKGV+D TA E +++
 Sbjct: 71 SLEDLDEVKALIDENEDKTLVQVDQTPAHLSSLEKDKGVIFTDILTAARESSDNLV 130
 Query: 112 EAHFGQ-ALAFDEDKLAAYHTAYFNSAAVLVDPHLEITTFIRAIPLQSDSDVPFNGHV 170
 E +F + + DE KL A H A N A LYVP +++ TP+AA+++ +S+ FN HV
 Sbjct: 131 EKYPMKDGKVKVDEHKLTALHAALVNGGAFVVPKNVQVETPQAVYVHESNDTALFN-HV 189
 Query: 171 LVIAQKESKFTYLERFESIGNATQKISANISVEVIAQGSQIKFSAIDRLGSPVTTYISR 230
 L+A S TY+E + S N + NI EVI + + + A+D L VTTY+R
 Sbjct: 190 LIVAEDHSSVTYVENYISTVNPKNVFN-NTISEVITGDNASVTYGAVDNLSGVTTYNKR 248
 Query: 231 RGRLE-KDANIDWALAVNENGVADPDSDLIQGSQADLKVVAASSGROVQGDITRVIN 289
 RG +D+ I+WAL +MN+G+ I++ ++L G G+ D K V G Q + T++ +
 Sbjct: 249 RGAARGRSKIEWALGNNDGTTISENTNLYGDTGYGDTKTVVVGROBQNTFTYLIH 308
 Query: 290 YQQRVGHILQHGVLIEROTLTFNGIGHILKDAKGADAAQSRVLMISDQARADANPIL 349
 +G+ + G+IL+HG+V+ + + FNGIG I A A++QSRVLMIS++AR DANPIL
 Sbjct: 209 FQKASRGYILKHGVKMSDASSITFNGIKTEHGASKANAQSRVLMISEKARGDANFIL 368
 Query: 350 IDENEVTAGHAASIGQVDPEDMYJMSRGLDQETAERLVIRGFLGAVIARIPIPSVRQE 409
 IDE+VTAGHAAS+G+VDP ++YLMERG+ +E AERLVI GFL V+ +PI V+++
 Sbjct: 369 IDEDDVTAGHAASVGRVDPIDLYYJMSRGIPEKRAERLVITGYFLAPVWNLPIEGVKQL 428
 Query: 410 IKVLDEKL 417
 + V++ K+
 Sbjct: 429 VSVIERKV 436

An alignment of the GAS and GBS proteins is shown below.

Identities = 322/420 (76%), Positives = 368/420 (86%)

Query: 1 MSKEAILNFLQARGEPTWLQELRLKAPKIELELFPVIERVKFHRWNLGDGTILENDYTA 60
 N-KS ++ F QA EP WLQE RL A E I LELP IERVKFHRWNLGDGT+ EN+ A

-2100-

Sbjct: 1 MTKKKLWAPSQAFAPAMLQERRLAALAEIPNLKLPPTIERVKFHRWNLGDGTLTENESLA 60

Query: 61 NVPDFTELGNPKLVQIGTQTVLEQVPMELIEKGVVTFDFSALEEIPEVIERYFGKARP 120
+VPDF +G+NPKLVQ+GTCTVLEQ+PM LI+KGVVF+DFY+ALEEIPVIE +FG+A

5 Sbjct: 61 SVPDFIAGDENPKLVQVGTQTVLEQLPMALIDKGVVDFPFTYALEEIPVIEAHFGQALA 120

Query: 121 FEEDRLAAHYHTAYFNSGAVLYIPDNVEITQPIEGLFYQDSQSKVPFNKHILLIVGNKAV 180
F+ED+LAAHYHTAYFNS AVLY+PD++EIT PIE +F QDS S VPPNKH+L+I GK +K

10 Sbjct: 121 FDEDKLAAHYHTAYFNSAAVLYVFDHLEITTPIEAIFLQDSQSDVPFNKHVLIAGKSKF 180

Query: 181 SYLERFESIGDGTERTSANISVEVIAQAGSQIKFASIDRIAGNVTTFISRRGRHSSDATI 240
+YLERFESIG+ T++ SANISVEVIAQAGSQIKF++IDRI +VT+ISRRGR DA I

Sbjct: 181 TYLERFESIGNATQKISANISVEVIAQAGSQIKFASIDRIAGPSVTTFISRRGRLEKDANI 240

15 Query: 241 DWALVMNEGNVADFDSDLTIGDSHANLKVVAASSGRVQGIDTRVTNYGCNSVGHILQ 300
DWAL VMNEGNV+ADFDSDLTIG GS A+LKVVAASSGRVQGIDTRVTNYG +VGHILQ

Sbjct: 241 DWALVMNEGNVIADFDSDLTIGGQSQADLKVVAASSGRVQGIDTRVTNYGQTVGHILQ 300

Query: 301 HGVILERGTLTFNGIGHIKAKGADRAQQESRVMLSDKASDANPILLIDENDVTAGHA 360
HGVILERGTLTFNGIGHI+K AKGADRAQQESRVMLSD+AR+DANPILLIDEN+VTAGHA

20 Sbjct: 301 HGVILERGTLTFNGIGHILKADKADRAQQESRVMLSDQAKADANPILLIDENVTAGHA 360

Query: 361 ASIGQVDPEDLYLMSRGLNQKTABQLVIRGFLGTVIAELPVKEVRDEMIADVTIKLEKR 420
ASIGQVDPED+YYLMSRGL+Q+TAE+LVIRGFLG VIAELP+ VR E+I V+D KL R

25 Sbjct: 361 ASIGQVDPEDMYLMSRGLDQCTAEERLVIRGFLGAVIAELPIPSVQEIIVKDEKLLNR 420

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1861

30 A DNA sequence (GBSx1968) was identified in *S. agalactiae* <SEQ ID 5785> which encodes the amino acid sequence <SEQ ID 5786>. This protein is predicted to be ABC transporter, ATP-binding protein, Ycf16 family. Analysis of this protein sequence reveals the following:

Possible site: 59
>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----

bacterial cytoplasm ---	Certainty=0.2253 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

40

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15260 GB:299120 similar to ABC transporter (ATP-binding protein) [Bacillus subtilis]

Identities = 180/250 (72%), Positives = 212/250 (84%)

45 Query: 2 SVLEIKNLHVSIEDKEIKGLNLTLKTGEIAAIMSPNGTGKSTLSAALINGPNYEVYTAG 61
S L IK+LHV IE KEILG+NL +K GE A+MSPNGTGKSTLSAALIM+P YEVY G

Sbjct: 4 SILTIKDLHVEIEBKILKGLNLEIKGGEFHAVMSPNGTGKSTLSAALIMGHPCYEVYTKS 63

50 Query: 62 ILFDGEDILEVDERARLGLFLAMQYPSEVPGITNARFIRAAAMAGKADDDKISIRQFI 121
I DG+D+LE+EVDERA+ GLFLAMQYPSE+ G+TNA+F+R+A+NA + + D+IS+ +FI

Sbjct: 64 IITLGGKDLVEMRVDRAQGLFLAMQYPSEISGVTNADFLRSAINARREGDEISIMKFI 123

55 Query: 122 TKLDERNELLGMKREMAERYLNEGFSGGEKKKNNHILQLMLEPKFALLDEIDSGLDIDL 181
K+DE MR L M EMA+RYLNEGFSGGEKKKNNHILQL+M+EPK A+LDEIDSGLDIDL

Sbjct: 124 RMIDENNEFLMDPEMAQRYLNEGFSGGEKKKNNHILQLMLEPKFALLDEIDSGLDIDL 183

Query: 182 KVVSGVNEMRGSGFGAMITTHYQRLNLYITPDKVHVMDCKVLSGGPELAVRLEKEGY 241
KVVSGH+N+MR E FG ++ITHYQRLNLYITPD VHVMM G+VV SGG RLA RLE EGY

60 Sbjct: 184 KVVSGINKMRSENFGLMITTHYQRLNLYITPDVHVMMGGRVVSQSGARLARLEAEGY 243

-2101-

Query: 242 AQIAREELGLR 251
 I +ELG+R
 Sbjct: 244 DWIKQELGIR 253

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5787> which encodes the amino acid sequence <SEQ ID 5788>. Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2417 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 An alignment of the GAS and GBS proteins is shown below.

Identities = 225/255 (88%), Positives = 241/255 (94%)

Query: 1 MSVLIEIKNLHVSIEDKEILKGINLITAKTGEIAAIMGPNCTGKSTLSAAIMGNPNYEVTAQ 60
 MS+LEI NLHVSIE KEILKG+NLITKTGE+AAIMGPNCTGKSTLSAAIMGNPNYEV T G
 Sbjct: 1 MSILIEINLHVSISIGKEILKGINLITAKTGEVAAIMGPNCTGKSTLSAAIMGNPNYEVTAQ 60

Query: 61 EILFDGEDILELEVDERARLGLFLAWQYPSEVPGITNAEFRAAMNAGKADDDKISIRQF 120
 +IL DG +IL+LEVDERARLGLFLAWQYPS+PGITNAEF+RAAMNAGKAD+DKIS+R F
 Sbjct: 61 QILLDGVNILDLEVDERARLGLFLAWQYPS+PGITNAEFRAAMNAGKADDDKISVDF 120

Query: 121 ITKLDERKELLGMKEEMAEYRLNEGFSGGKKRNEILLQMLLEPKFALLDEIDSGLDIDA 180
 ITKLDERK LLGMKEEMAEYRLNEGFSGGKKRNEILLQMLLEPKFALLDEIDSGLDIDA
 Sbjct: 121 ITKLDERKALLGMKEEMAEYRLNEGFSGGKKRNEILLQMLLEPKFALLDEIDSGLDIDA 180

Query: 181 LKVSFKGVNENRGSGPGAMIIITHYQRLNLYITPDVKVHVNMDGKVLVSGGPELAVLEKEG 240
 LKVSFKGVNENRG+ PGAMIIITHYQRLNLYITPD VHVNMKG++VLGG LA RLEKEG
 Sbjct: 181 LKVSFKGVNENRGSGPGAMIIITHYQRLNLYITPDVHVNMMDGRIVLSGDAALATRLKEG 240

Query: 241 YAQIAREELGLEVEE 255
 YA IA++LG+EYKEE
 Sbjct: 241 YAGIAQDLGIEYKEE 255

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 1862

A DNA sequence (GBSx1969) was identified in *S.agalactiae* <SEQ ID 5789> which encodes the amino acid sequence <SEQ ID 5790>. This protein is predicted to be RgpG (rfe). Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -12.10	Transmembrane	312 - 328 (308 - 336)
INTEGRAL	Likelihood = -10.03	Transmembrane	15 - 31 (6 - 41)
INTEGRAL	Likelihood = -9.82	Transmembrane	205 - 221 (197 - 226)
INTEGRAL	Likelihood = -8.60	Transmembrane	335 - 351 (329 - 358)
INTEGRAL	Likelihood = -7.48	Transmembrane	257 - 273 (255 - 281)
INTEGRAL	Likelihood = -5.52	Transmembrane	60 - 76 (56 - 79)
INTEGRAL	Likelihood = -5.31	Transmembrane	151 - 167 (148 - 171)
INTEGRAL	Likelihood = -4.88	Transmembrane	91 - 107 (90 - 108)
INTEGRAL	Likelihood = -4.78	Transmembrane	184 - 200 (177 - 203)
INTEGRAL	Likelihood = -3.13	Transmembrane	119 - 135 (119 - 135)
INTEGRAL	Likelihood = -2.97	Transmembrane	229 - 245 (229 - 250)

----- Final Results -----
 bacterial membrane --- Certainty=0.5840 (Affirmative) < succ>

-2102-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8919> which encodes amino acid sequence <SEQ ID 8920>

5 was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
McG: Discrim Score: 5.18
GVH: Signal Score (-7.5): -6.19
Possible site: 15
10 >>> Seems to have an uncleavable N-term signal seq
ALOM program count: 9 value: -12.10 threshold: 0.0
INTEGRAL Likelihood = -12.10 Transmembrane 239 - 255 (235 - 263)
INTEGRAL Likelihood = -9.82 Transmembrane 132 - 148 (124 - 153)
15 INTEGRAL Likelihood = -8.60 Transmembrane 262 - 278 (256 - 285)
INTEGRAL Likelihood = -7.48 Transmembrane 184 - 200 (182 - 208)
INTEGRAL Likelihood = -5.31 Transmembrane 78 - 94 (75 - 98)
INTEGRAL Likelihood = -4.88 Transmembrane 18 - 34 (17 - 35)
INTEGRAL Likelihood = -4.78 Transmembrane 111 - 127 (104 - 130)
20 INTEGRAL Likelihood = -3.13 Transmembrane 46 - 62 (46 - 62)
INTEGRAL Likelihood = -2.97 Transmembrane 156 - 172 (156 - 177)
PERIPHERAL Likelihood = 12.63 284
modified ALOM score: 2.92

*** Reasoning Step: 3

----- Final Results -----
bacterial membrane --- Certainty=0.5840 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
30 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA82114 GB:AB022909 RgpG [Streptococcus mutans]
Identities = 266/382 (69%), Positives = 317/382 (82%)
35 Query: 10 TIRYIFVLIGAFLLSIILTPIRVISLKVGVADKPNARRINKVMPSPGGGLAIFLSEFVTI 69
T+++ VLI L S++LTP++R +L+VGAVD PNARRINKVMPSP+GGGLAI+SFV+
Sbjct: 7 TLKFLVLIATLTLTSLVLTPLRVFPALRVGAVDNPARRINKVMPSPAGGLAIIISFVIA 66
40 Query: 70 TLFPMPMAASRHFIEVSFYHYLLPVIIGGLVTTTGTGIDDFELRPRYRMIGIIIAAIII 129
TL MPM SYF YILPV++G LV+ TGFID++EL P+ K LGI++ A+II
Sbjct: 67 TLALMPMLIKTQIQGKSYFEYLLPVVLGALVIALTGFIDVYELSPKIKFLGILLGAVII 126
45 Query: 130 WKFTAPRDFSPKIPIGGFLLEPGVILTFPLTVLWIIISITNAINLIDGLDGLVSGVSIISL 189
W FT PRFDSPKIP GGP+L F P L+PFLT+LW+++ITN+NLIDGLDGLVSGVS+ISL
Sbjct: 127 WIPTDFRFDSPKIPPGGFLHFNFLSFFLTILWVAITNAVNLIDGLDGLVSGVSMISL 186
50 Query: 190 ATGAVVSYFFLPKIDFPLTITIVILASIVGFFPNYHYPAIIYLGDAAGALFIFGMIVLS 249
TM +VSYYFL D PLTITI +LI +I GFFPNYHYPAIIYIGD GALFIFGMIVLS
Sbjct: 187 TMGLVSYFFLYDITDITPLTITITFVLPAIAGFFPNYHYPAIIYIGDITGALFIFGMIVLS 246
55 Query: 250 LQGLGNSTAVAVITPVIIIGVPIIDTAVAIVRKLSGKKIISADRMHLLHRLLSMGFTIR 309
LQGLGN+TAVAV+TP+I+LGVP+IDT VAI+RR LSG+K EAD MHLHRL+MGFTIR
Sbjct: 247 LQGLGNATAVAVVTP+IIVLGVP+VDTTVAIIRRTLSGKFFYRAIDNMHLLHRLLSMGFTIR 306
60 Query: 310 GAVLVVYGIATIFSLIALLNVSSRIGGIFILLALLAMEIFTEGLNIWGENRTPLENL 369
GAVLVVYGIAT+ FSL++LLNVSSR+GGI L++ + A+EIFEEL IWG KTLPL LL
Sbjct: 307 GAVLVVYGIAMFFSLVSLLLNVSSRLGGILLMIGVAFALFIEGLEIWSGPKRTPLFRLL 366
Query: 370 KFIGNSDYRQSVIAKYSKHKQK 391
FIGNSDYRQ V+AKY K+K
Sbjct: 367 AFIGNSDYRQEVVAYRRKKKK 388

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5791> which encodes the amino acid sequence <SEQ ID 5792>. Analysis of this protein sequence reveals the following:

-2103-

Possible site: 32
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -8.28	Transmembrane	9 - 25 (1 - 33)
INTEGRAL	Likelihood = -8.17	Transmembrane	201 - 217 (198 - 221)
INTEGRAL	Likelihood = -7.64	Transmembrane	308 - 324 (305 - 329)
INTEGRAL	Likelihood = -7.17	Transmembrane	55 - 71 (51 - 74)
INTEGRAL	Likelihood = -7.06	Transmembrane	145 - 161 (138 - 170)
INTEGRAL	Likelihood = -6.58	Transmembrane	260 - 276 (251 - 278)
INTEGRAL	Likelihood = -6.21	Transmembrane	180 - 196 (172 - 198)
INTEGRAL	Likelihood = -5.95	Transmembrane	331 - 347 (330 - 353)
INTEGRAL	Likelihood = -5.68	Transmembrane	87 - 103 (82 - 104)
INTEGRAL	Likelihood = -3.93	Transmembrane	113 - 129 (112 - 133)
INTEGRAL	Likelihood = -2.60	Transmembrane	233 - 249 (232 - 250)

----- Final Results -----
 bacterial membrane --- Certainty=0.4312 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAA82114 GB:AB022909 RgpG [Streptococcus mutans]
 Identities = 289/381 (75%), Positives = 334/381 (86%), Gaps = 1/381 (0%)

Query: 5 TIDYVLVLIGALINSLFLTPLVRFALFRVGVADNPARRRNVKVPMPSSGLLAIFMSPLVA 64
 T+ +VLVLI L L SL LTPLVRF A RVGVADNPARRRNVKVPMP+GGIAI +SP+AA
 25 Sbjct: 7 TLKFPVLVLTATLTLTLVLTPLVRFALFRVGVADNPARRRNVKVPMPSSGLLAIFMSPLVA 66

Query: 65 SLGLPIASKGMFFGQTYFSYILEPVVIGATVITLTGFLDDLYELSPKLMFGILIGAVI 124
 +L L+P+ K G++YF YILEPV+GA VI LTGF+DD+YELSPK+K GIL+GAVI
 30 Sbjct: 67 TLALMPMLK-TQGGKSYFEVILEPVVLGALVIALTGFDVYELSPKLPGLLIGAVI 125

Query: 125 VWAFITDFKFSFKIPFGGPLLIVFGPFLTLFLTLVLMIVSITNAINLIDGLDGLVSGVSIIS 184
 +W FTDF+FDSPKIPFGGP+L F PFL+ FLT+LM+V+ITNA+NLIDGLDGLVSGVSIIS
 35 Sbjct: 126 IWIPTDFRDFSPKIPFGGPMLEFNPFLSFFLTLVLMVLTNAINLIDGLDGLVSGVSIIS 185

Query: 185 LVTMAIVSYFFLPQKDFLTLTLVLISALAGFFPNYHPEAMITLGTGALFGIMIGVL 244
 L TM +VSYPFL D FLTLTL VLI ALAGFFPNYHPE+IYLGTGALFGIMIGVL
 40 Sbjct: 186 LITMGLVSYFFLMDTDFLTLTLTLVLIFALAGFFPNYHPEAMITLGTGALFGIMIGVL 245

Query: 245 SLQGLKNSTAVAVVTPVILGVPIMDTIVAIRRLSQQKPYEDKMLHRLHLLSMGFTH 304
 SLQGLKN+TAVAVVTP+I+LGVPI+DT VAIIRR+LSQQKPYE D MHLHRLHLL+MGFTH
 40 Sbjct: 246 SLQGLKNSTAVAVVTPVILGVPIMDTIVAIRRLSQQKPYEDKMLHRLHLLSMGFTH 305

Query: 305 RGAVLVYVYGITMLPSLISLLNVSSRIQGVLLMLGLLPGLEVFIHGLEWGEKRTPLFNL 364
 RGAVLVYVYGI M PSL+SLLLNVSSR+GG+LLM+G+ F LE+FIHGLEWGEKRTPLFNL
 45 Sbjct: 306 RGAVLVYVYGIAMFSLISLLNVSSRLGILLMGVAFLEIFIHGLEWGEKRTPLFNL 365

Query: 365 LKFIGNSDYRQAMLLWKKEK 385
 L FIGNSDYRQ ++ K++ KK
 50 Sbjct: 366 LAFIGNSDYRQEVVAKYRRKK 386

An alignment of the GAS and GBS proteins is shown below.

Identities = 282/384 (73%), Positives = 334/384 (86%), Gaps = 1/384 (0%)

Query: 6 MIPPTIEYIFVLIGAFLLSIIITPIIRVISLKVGVADKPNARRRNVKVPMPSSGLLAIFLS 65
 M FTL+Y+ VLIGA L+S+ LTP+R ++ +VGVD PNARRRNVKVPMP+SGGLAIF+S
 55 Sbjct: 1 MFSPTIDYVLVLIGALINSLFLTPLVRFALFRVGVADNPARRRNVKVPMPSSGLLAIFLS 60

Query: 66 FVVTTLFMPMAAS-RHFIEVSYHFIYLPVILGGVLTITGFDIDFELPRPYKGLGII 124
 F+V +L +P+AA+ F +YF YILEPV+IG V+T TGF+DD+REL P+ KM GI+I
 60 Sbjct: 61 FLVASLGLPIASKGMFFGQTYFSYILEPVVIGATVITLTGFLDDLYELSPKLMFGILIGAVI 120

Query: 125 AAILIWKFTHFRFDSFKIPFGGPLLIFGPIILTFVLVMIISITNAINLIDGLDGLVSGV 184
 A+I+W FT F+FDSPKIP GGPLL FGP LT FVLVMI+STNAINLIDGLDGLVSGV
 65 Sbjct: 121 GAVINVAFTDFKFSFKIPFGGPLLIVFGPFLTLFLTLVLMIVSITNAINLIDGLDGLVSGV 180

-2104-

Query: 185 SIISLATMAVVSFFFLPKIDFFILTLITVILIASIVGFFFPYNYHPIIYLGDAAGALFTGM 244
 SIISL TMA+VSFFFLP+ DFFILTLTI++LI++I GFFFPYNYHPI+IYLG DAAGALFTGM
 Sbjct: 181 SIISLVTMAIVVSFFFLPKQDFFILTLITVILIASIAGFFFPYNYHPAMIYLGDTGALFTGM 240

5 Query: 245 IGVLSLQGLKNSTAVAVITPVIIIGVPIITAVAIIVRKLSSGKKISBEADKMLHRLHLLSM 304
 IGVLSLQGLKNSTAVAV+TPVIIIGVPI+DT VAI+RR LSG+K R DKMLHRLHLLSM
 Sbjct: 241 IGVLSLQGLKNSTAVAVITPVIIIGVPIITAVAIIRRSLSGQKRYEPDKMLHRLHLLSM 300

10 Query: 305 GPTHRGAVLVVYGIATPSLIALLNVSSRIGGIFLLIALLAMKIFTEGLINTWGENRTP 364
 GPTHRGAVLVVYGI ++PSLI+LLLNVSRRIGG+ L+L LL +E+PIEGL INGE RTP
 Sbjct: 301 GPTHRGAVLVVYGITMPSLSIALLNVSSRIGGVLLMLGLLPLEVFTIEGLETWGEKRTTP 360

Query: 365 LFNLLKFTIGNSDYRQSVIAKYSK 388
 LFNLLKFTIGNSDYRQ+++ K+ +K
 15 Sbjct: 361 LFNLLKFTIGNSDYRQAMLLKWEK 384

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1863

- 20 A DNA sequence (GBSx1970) was identified in *S. agalactiae* <SEQ ID 5793> which encodes the amino acid sequence <SEQ ID 5794>. This protein is predicted to be negative regulator of genetic competence. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3460 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30

A related GBS nucleic acid sequence <SEQ ID 9483> which encodes amino acid sequence <SEQ ID 9484> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

- >GP:BAAS2113 GB:AB022909 negative regulator of genetic competence
 [Streptococcus mutans]
 35 Identities = 168/248 (67%), Positives = 205/248 (81%), Gaps = 9/248 (3%)
- Query: 1 MEMKQISSETTLKITTISMRDLERGMELKDFLIPOKTEERFFPYVMDELLDPENFNSGML 60
 MEMKQISSETTLKITTISMRDLERGMELKDFLIPOKTEERFFPYVMDELLDPENFNSGML
 40 Sbjct: 1 MEMKQISSETTLKITTISMRDLERGMELKDFLIPOKTEERFFPYVMDELLDPENFNSGML 60
- Query: 61 SFRVTPKDRIDVFFVTKSELKDIANLEADIDISKMSPEDFFKTLQSMLEKEDTDAH 120
 SFRVTP+ DRIDVFFVTKSE++K+LNLSE+L+D DISKMSPEDFF+TLE++H+EKGD A
 45 Sbjct: 61 SFRVTFNRDRIDVFFVTKSEINKNLNLEDLDFDDISKMSPEDFFFTLE++H+EKGDAAAL 120
- Query: 121 AKLAKIKNMMDKATCEVEEENVSSEQPEKEVEITIGYVHVVDPMNIRKAVVPSQTIDFPI 180
 KLAETIE ++TQ+ E+ ++E+ + YVHV+V DE NI+ V+ E++TID+ +
 Sbjct: 121 DKLAETIEKREKTKC--EKGETKEKRD-----YVHVLDPFNIQQVTSFAKTVVDIV 171
- 50 Query: 181 EASELYKNGKGYHMTILLDLNQPSTYFANLMYARMLEHNAVCTKTRAYLKESHLIHD 240
 EASEL+K YHMT+LL+LE++P Y+AM+ARMLEHA CTKTRAYL EH +CLI D
 Sbjct: 172 EASELPKESDAYHMTVILNLEDKPQYYADLMFARMLEHAGRGTKTRAYLLSHGVQLIKAD 231
- Query: 241 AISKLMQI 248
 A+ +LQMI
 55 Sbjct: 232 ALQKLMQI 239

-2105-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5795> which encodes the amino acid sequence <SEQ ID 5796>. Analysis of this protein sequence reveals the following:

```

Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3307 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 171/253 (67%), Positives = 209/253 (82%), Gaps = 2/253 (0%)

Query: 1  MEMKQISSETLLKITISMEDLEDRGMELKDFLIPQEKTEFFYSVMDELDLPFNFKNSGML 60
Sbjct: 3  MEMKQISSETLLKITISM+DLE+RGMELKDFLIPQEKTEFFYSVMDELDLP+NFK+SGML 62

Query: 61  SFRVTPKDRIDVFTKSELSKDINLEELADLQDISKMSPEDFKTLQSKMSKGTDRH 120
Sbjct: 63  SFRVTPKDR+KDR+DVFTKSE++KD+NLE+LA+ GD+S+H+PEDFTK+LEQSM EKGD AH 122

Query: 121 AKLAEINEMDKATQEVV--EENVSEQFEKEVETIGYHYVDFDNIEAVVRPQCTIDF 178
      XL +IE +K+ + + + ++ E E + YHYV DF I V F++TIDF
Sbjct: 123 EKLEKIEIMEDVVEATLANQSEADPSTNHESEPLDTYHYVLDFTITAEVAFKTDIF 182

Query: 179 FIEASELYKNGKGTHTMTLLDLENQPSYFANLMYARMLEHANGTKTRAYLKHSIQLIH 238
      IEASELYK YHMTLLD++ QPSYFAN+MYAR++ERAN G+KTRAYL+EH +QI+
Sbjct: 183 SIEASELYKNGSNTYHMTLLDVQQQPSYFANVMYARLIEHANGKTRAYLQEHGLQML 242

Query: 239 DDAISKLMIEWG 251
      D A +LQ IE+G
Sbjct: 243 DGAVEQLKIELG 255

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1864

A DNA sequence (GBSx1971) was identified in *S.galactiae* <SEQ ID 5797> which encodes the amino acid sequence <SEQ ID 5798>. This protein is predicted to be BacA (bacA). Analysis of this protein sequence reveals the following:

```

Possible site: 17
>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -9.02 Transmembrane 115 - 131 ( 111 - 135)
INTEGRAL Likelihood = -8.97 Transmembrane 227 - 243 ( 219 - 247)
INTEGRAL Likelihood = -7.86 Transmembrane 48 - 64 ( 44 - 69)
INTEGRAL Likelihood = -7.27 Transmembrane 263 - 279 ( 260 - 279)
INTEGRAL Likelihood = -7.22 Transmembrane 87 - 103 ( 85 - 107)
INTEGRAL Likelihood = -3.50 Transmembrane 2 - 18 ( 1 - 19)

----- Final Results -----
      bacterial membrane --- Certainty=0.4609 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AA050462 GB:AF169967 BacA [Flavobacterium johnsoniae]
Identities = 101/275 (36%), Positives = 165/275 (59%), Gaps = 22/275 (8%)

Query: 7  LKALFLGVVSGVTEWLFVSSSTGHLLVLRQFMKLNQSKSFVEMFNIVQLGAIMAVIVYIF 66

```

-2106-

L+A+ L V+EG+TE+LPVSTGH+I+ F + + F ++F IVIQLGAI++V++YF
 Sbjct: 4 LQAIVLAVIRGITFLPVSTGHMILASSFPGIAH-EDPTKLTFTVIQLGAILGVVLYF 62
 Query: 67 KRINFPQPGKRSRRIRLTWQLWKVVIACIPISILIAFPDNNFRAHFNPMIPAIALIPY 126
 KR PQ T + K+++A IP++++ L ++ + +A++L+
 Sbjct: 63 KRF--PQ-----TLDFTFKLLVAFIPAVVLGILLSDFTDGLLENRPVTVAVSLIG 110
 Query: 127 GFVFI---WVEKRNHLKPKQVTEIASMSYKTAFLIGCPQLSIVPGTSRSGATILGAI 182
 G + + W NA Q ++Y A IG PQ +++PG SRSGA+I+G +
 Sbjct: 111 GLILLKVDWPNPNNAASTSQ-----KITLYQALKIGLPGCIAMIPGVSRSGAIVGMS 165
 Query: 183 IGTSSVAADFTFFLAIPMTFGYSLKAVKVFLDGNVLSLDQSLILLVASLTAFFVSLNV 242
 SR+ AA+F+FFLA+PIM G + K Y+ G LS DQ IL++ ++ AP+V+L
 Sbjct: 166 QKLKSTTAARFSPFLAVPTMIGATVKKCYDYKAGFELSHDQVNLIIIGNVAFIVALLA 225
 Query: 243 IRFLTDYVKRHDFITFGKYRIVLGSLLIYLWLVH 277
 I+ ++ ++ F +PG YRI+ G +L+L +H
 Sbjct: 226 IKTFISFLTKNGFKVFGYRIAGIILLIHFFIH 260

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5799> which encodes the amino acid sequence <SEQ ID 5800>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence
 25 INTEGRAL Likelihood = -11.30 Transmembrane 225 - 241 (219 - 247)
 INTEGRAL Likelihood = -9.24 Transmembrane 115 - 131 (109 - 135)
 INTEGRAL Likelihood = -7.64 Transmembrane 48 - 64 (44 - 69)
 INTEGRAL Likelihood = -7.43 Transmembrane 87 - 103 (85 - 108)
 INTEGRAL Likelihood = -5.15 Transmembrane 263 - 279 (262 - 279)
 30 INTEGRAL Likelihood = -3.82 Transmembrane 2 - 18 (1 - 19)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAD50462 GB:AF169967 BacA [Flavobacterium johnsoniae]
 Identities = 102/269 (37%), Positives = 169/269 (61%), Gaps = 14/269 (5%)
 40 Query: 7 LKAIFPGIRGITWLPVSSTGHLLVQCFIRINQDKAFIRMFNIVIQIGAIIVMLIYF 66
 L+AI +IBGITE+LPVSSTGH+I+ F + + F ++F IVIQLGAI++V++YF
 Sbjct: 4 LQAIVLAVIRGITFLPVSTGHMILASSFPGIAH-FTKLTFTVIQLGAILGVVLYF 62
 Query: 67 KRINFPQPGKRSRRIRLTWQLWKVVIACIPISILIAFPDNNFRAHFNPMIPAIALIVY 126
 +R FQ T + K+++A IP++++ L ++ + + V ++L+L+
 45 Sbjct: 63 KRF--PQ-----TLDFTFKLLVAFIPAVVLGILLSDFTDGLLENRPVTVAVSLIG 110
 Query: 127 GIAPWIKRKNQAQRPVTEIARMSYKTAFFIGCPQLSIVPGTSRSGATILGAILGTS 186
 G+ + +++ A T +++Y A IG PQ +++PG SRSGA+I+G + S
 50 Sbjct: 111 GLILLKVDWPNPNNAASTS-OKITLYQALKIGLPGCIAMIPGVSRSGAIVGMSQKLS 169
 Query: 187 RTVAADFTFFLAIPMTFGYSLKAVKVFLDGHFLDQVILLVASLTAFFVSLAIRFL 246
 RT AA+F+FFLA+PIM G + K ++ G L QV IL++ ++ AP+V+LAI+
 Sbjct: 170 KTTAAEFSPFLAVPTMIGATVKKCYDYKAGFELSHDQVNLIIIGNVAFIVALLAIF 229
 55 Query: 247 TDYVKHDFITFGKYRIVLGSLLIYSFF 275
 ++ K+ F+PG YRI+ G +L+L FF
 Sbjct: 230 ISFLTKNGFKVFGYRIAGIILLIHFF 258

60 An alignment of the GAS and GBS proteins is shown below.

Identities = 227/272 (83%), Positives = 253/272 (92%)

Query: 1 MLIIELKALPIGVVEGVTEWLPVSSTGHLLVQCFIRINQDKAFIRMFNIVIQIGAINA 60
 MLIIELKKA+P G++EG+TEWLPVSSTGHLLVQCF++INQ K+F+RMFNVIVQLGAI+A

-2107-

Sbjct: 1 MLIIELLKAIFPGIIGITWELPVSSTGHLILVGEFIRLQDKAFIEMFNIVIQGLAII 60
 Query: 61 VIVIVPKRLNPPQPGKAREIRLITWQLNKVVIACIPSLIALPFDNMFRAHFNMPDIA 120
 V++IYP+RLNPPQPGK+ARE++ITWQLNKVVIACIPSLIA+P DNMFRAHF F++PIA
 5 Sbjct: 61 VMLIYPERLNPPQPGKAREIRLITWQLNKVVIACIPSLIAVPLDNMFRAHFTFMPDIA 120
 Query: 121 IALIFYGFVFIWVEKRNHKLKQVTELASMSYKTAFLICGQVLSIVPGTSRSGATILGA 180
 IALI YG FIW+EKRNA +P VTEIA MSYKTAFL ICQCQVLSIVPGTSRSGATILGA
 Sbjct: 121 IALIVGIAFIWIEKRNACQBPVATLARMYSYKTAFFICGQVLSIVPGTSRSGATILGA 180
 10 Query: 181 IITGTSRVAADFTFLAIFTMFGYSGLKAVKFLDGNVLSLDQSLILLVASLTAFVVSIL 240
 II+GTSR+VAADFTFLAIFTMFGYSGLKAVK+FLDG+ L Q LILLVASLTAFVVSIL
 Sbjct: 181 IILGTSRVAADFTFLAIFTMFGYSGLKAVKFLDGHLLDFAQVILLVASLTAFVVSIL 240
 15 Query: 241 YVIRFLADYVKRHDFITFGKYRIVIGSLLILY 272
 IRFLADYVK+HDFITFGKYRIVIGSLL++Y
 Sbjct: 241 LAIRFLADYVKRHDFITFGKYRIVIGSLLILY 272

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1865

A DNA sequence (GBSx1972) was identified in *S. agalactiae* <SEQ ID 5801> which encodes the amino acid sequence <SEQ ID 5802>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence
 25 INTEGRAL Likelihood = -8.65 Transmembrane 494 - 510 (488 - 519)
 INTEGRAL Likelihood = -8.01 Transmembrane 263 - 279 (256 - 288)
 INTEGRAL Likelihood = -5.95 Transmembrane 25 - 41 (20 - 43)
 30 INTEGRAL Likelihood = -4.94 Transmembrane 475 - 491 (473 - 493)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9481> which encodes amino acid sequence <SEQ ID 9482> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA899606 GB:U67598 M. jannaschii predicted coding region MJ1577
 40 [Methanococcus jannaschii]
 Identities = 41/172 (23%), Positives = 78/172 (44%), Gaps = 19/172 (11%)
 Query: 479 LISFVLIITFLINYPFTYFCIYLLLVGILLINKIIPMTKRKISNGYIVTEGASRVYQW 538
 +IS ++ ++ F+ ++ + +++ II +T G ++ +W
 45 Sbjct: 442 VISILLAVFLYPIPKYSQTFNEVFLSIVFVVQNTIILATITELPGRWKANYYEKL-EW 500
 Query: 539 TSFRMLRDIKSPDRSELESIVLNNKILVYATLFGYADRVKALR-VNQIIDIPERFANID 597
 +F+N L ++ + E I +W L+Y T G D+V +A++ +N ++ + I
 Sbjct: 501 DAFKNFLSLNIAIKYKSPEDISWKNLITCTALGVGDKVVEAKMSLNLSLNAVDTVIIH 560
 50 Query: 598 SHQFAISVNGSSNIPSTITEDVSHSNFSVNSGGSSGQFSGGGG--GGGGGA 647
 S+ ++ + S + ST GS GGF GGG GGGGGA
 Sbjct: 561 SNYDEMKTSVDVYSSTT-----GGGGGFGAGGGGFGGGGGA 597

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 5803> which encodes the amino acid sequence <SEQ ID 5804>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have a cleavable N-term signal seq.

-2108-

INTEGRAL Likelihood = -7.91 Transmembrane 486 - 502 (483 - 508)
INTEGRAL Likelihood = -5.89 Transmembrane 465 - 481 (460 - 483)
INTEGRAL Likelihood = -2.18 Transmembrane 244 - 260 (241 - 260)

5 ----- Final Results -----
bacterial membrane --- Certainty=0.4163 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

10 The protein has homology with the following sequences in the databases:

>GF:AA899606 GB:U67598 M. jannaschii predicted coding region MJ1577
[Methanococcus jannaschii]
Identities = 59/263 (22%), Positives = 106/263 (39%), Gaps = 14/263 (5%)

15 Query: 369 FLDMAFGNKVTLFVDQLFSGYHYDADTIKOLKCTYKGGKLEQEVRSSEQVIMKMGASA 428
++ G K+ + L + Y+D +K L K K + E +S Q K+ K
Sbjct: 346 YIKIMGGKIEILKTDLNLDVYESDVMKFLMKYKSNVDFPEYIKSLAQKYSKDKLK 405

20 Query: 429 AITNNVISTIKKINLFDIYRQMPTA--EKRRKSNVSGQLCLLIISGLLIYLAIKESGL 486
+ + E K+ P ++ A ER + L + ++L L ++
Sbjct: 406 KLKD--BLDKIMEYPRYSKVVNAPLETRGKICIIALLVISILLAVLPYPIPKYSQTFN 462

Query: 487 ALTYLALMVLTMCLGFSYSLKLDQYKKGIGETPEGGVRLHQWQSPKMIIRDIDPEVAI 546
+ YL+++ + I L L +W +PKN + + + +
25 Sbjct: 463 EVFPLSIVFVVG---NILLALTFTSLFGRNKANYKKELEWDAFNFLENLAMIKKYSP 518

Query: 547 EGLVNVNRLVYATLFGYAKQVRYLVKVRHIALPEVYQAVRPGELSMVMVYATTFPVSSL 606
E +W L+Y T G KV +K ++ + V + Y + T V S +
Sbjct: 519 EDISINKDWLYGTALGVGDKVKEAMKSLNLS----ELVADYVIHNSINDKNTSVDVS 573

30 Query: 607 SSAATSSNFVSSGGGIGGGGG 629
S+TT S +GG GGGGG
Sbjct: 574 YSSTSGGGGFGAGGGGFGGGGG 596

35 An alignment of the GAS and GBS proteins is shown below.

Identities = 241/635 (37%), Positives = 372/635 (57%), Gaps = 18/635 (2%)

Query: 22 MKKCPALICLALSFPMVSVQADEVDYNI PHYEGNLTIRHNSADPTEKVTYQFDSYNGQ 81
MKK + + L S + ++A +VDY+I +YES L + +N+A F +KVTYQFD+SYNGQ
40 Sbjct: 1 MKKILMTLVLCFSLLGIRIKAADVDYGIITNYEGQILLKENTARFQKVTYQFITSYNGQ 60

Query: 82 VYTLGTAGLKEDNFNINNKQVEVINGKVRKVSQIEDLEDGYRLKVPNGGAGDITVKV 141
+Y+LG G LE F I+ KP+VEV NG+ VS + DL DGYRLK++N G+AGD V V
Sbjct: 61 YISLGRTHLEAGFAIDQKPKVEVYQNGQQVPVSQESDLGDRYRLKYNAGQAGKVDV 120

45 Query: 142 NVQNKLNVLFMHEDVGLNMTIPISDMDKTLEKVDWFISTDKKVALSLRHLGLYL-KTP 200
V W+L ++L ++DV ELNW PISDMDKTLEKV ++T + S LW H GY K P
Sbjct: 121 KVINQLHLLTAYQDVAELNMTIPISDMDKTLEKVSILVTVTPTDIDQSNLMARRGYQKPF 180

50 Query: 201 PKIRQNNKRYHLTAFNVNKKLEPHGYWDRSYF--NLFTSKNNKYKKIEYQSKMIRHGF 258
+++ N+RY + A NV+ +LE H YND+ P + + K KI E IR
Sbjct: 181 QVLKESNSRYQINAKVVSQGLHAYWDKALLGKEPVDVSTSKKKIVALETKISRRRT 240

Query: 259 ILSPLRLILLSPFFIIVTLFISIRVFLPRKKVNKYGQPKKHHLYEAPDLSPLELTQSI 318
+L L ++P + L+ I+ +K+ N+Y H YE PDLSP L LQ+ I
55 Sbjct: 241 LLQLLFGKVIPLVEVGFLMLQLQITRLKQPNRYHLANHTDHSYEVPELSPFLVLTQAI 300

Query: 319 YMSFKNFQ---DEEKKIHL---ISQELQISILLDLIDRKVL---NYDDNLLISLNL 368
Y SF E K + ++ E L+Q LLDLID+KVL L ++ LD
60 Sbjct: 301 YGQSFALSPITASQSKLLIFKGVTFALVQTLDDLIDQKVLALLKKEKAYLEISQLD 360

Query: 369 RASDARIDFIEFAPADGTSLKPDQLFSNYQFSYKETLRELKQKAKASDLQTMRRRGNA 428
R+D E F++ AF + +L DQLFS Y + +T+++LKK +K L+ ++R+
Sbjct: 361 RVTDEZAAFLDMARGNKVTLFVDQLFSGYHYD-AUTIKQLKCTYKGGKLEQEVRSSEQV 419

65 Query: 429 LSRITRLIRLISKDNINSLRRKGISSPYKMSSESSKSLRQFYSILPSIFVVIYT 488

-2109-

```

      + + + + I + + + + + YR+M+ E + + + + L + + + ++IY
Sbjct: 420 IKAMKASAAITNNVLETIKKLNLPDTRYQMTPAEKRRKSNVQGLGCLLLILNSGLLY- 478

5  Query: 489 LFLNYFTYPCIIYLLPGVILLNKKLIIPWTRKISNGYIVTEGASRVYQMTSFRNMLRDI 548
      L + + IYI L + + L I + + I T +G R++QW SF+NM+RDI
Sbjct: 479 LAIKESGLALIVLALMLVLMCLGPGYISLKLQYKKLGIEFPBGVRLHQWQSFKNMIRDI 538

Query: 549 KSFDRSELESIVLNMRLVYATLFGYADRVEKALKRVNQIDIPERFANIDSHQFAISVNCQ 608
      F + +E +V+MNR+LVYATLFGYA +VE+ L+V+I +PE + + + + + +
10 Sbjct: 539 DKFEDVATEGLVVMNRVLVYATLFGYAKKVERYLKVRIRIALPEVYQAVRPEGLSMVYAT 598

Query: 609 SNHFSITIEDVSHASINFSVNSGSGSGGFGGGGGG 643
      + F + + +SNFV+SG GG SGGGGGG
15 Sbjct: 599 TPTFVSSLSSATTSSNFVSSG---GGISGGGGGG 630

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 892> and protein <SEQ ID 892> were also identified. Analysis of this protein sequence reveals the following:

```

20 Lipop: Possible site: -1 Crend: 4
McG: Discrim Score: 10.29
GVH: Signal Score (-7.5): 3.11
      Possible site: 23
>>> Seems to have a cleavable N-term signal seq.
25 ALOW program count: 3 value: -8.65 threshold: 0.0
      INTEGRAL Likelihood = -8.65 Transmembrane 475 - 491 ( 469 - 500)
      INTEGRAL Likelihood = -8.01 Transmembrane 244 - 260 ( 237 - 269)
      INTEGRAL Likelihood = -4.94 Transmembrane 456 - 472 ( 454 - 474)
      PERIPHERAL Likelihood = 2.28 540
30 modified ALOW score: 2.23

*** Reasoning Step: 3

----- Final Results -----
35 bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no homology with any sequences in the databases.

40 Example 1866

A DNA sequence (GBSx1973) was identified in *S.galactiae* <SEQ ID 5805> which encodes the amino acid sequence <SEQ ID 5806>. This protein is predicted to be glutamine-binding periplasmic protein/glutamine transport system perme. Analysis of this protein sequence reveals the following:

```

45 Possible site: 24
>>> Seems to have a cleavable N-term signal seq.
      INTEGRAL Likelihood = -8.86 Transmembrane 301 - 317 ( 295 - 324)
      INTEGRAL Likelihood = -6.05 Transmembrane 479 - 495 ( 473 - 496)
      INTEGRAL Likelihood = -0.59 Transmembrane 369 - 385 ( 369 - 385)

50 ----- Final Results -----
      bacterial membrane --- Certainty=0.4545(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

55 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAA17584 GB:D90907 glutamine-binding periplasmic protein
      [Synechocystis sp.]
Identities = 147/534 (27%), Positives = 256/534 (47%), Gaps = 75/534 (14%)

```

-2110-

Query: 4 ILLSLPTALLITPGMTSIQADEYLRVGMGAAYAPFNWQNDNTNGAVPIEGIDQYANGY 63
 +Ll++ LL F ++ + + V E + P F T E T Q G+
 Sbct: 24 VLLAIALPLLPFAFSQVSR---CTIIVATEPTFPFFMTD-----EATQGLT-GF 68

Query: 64 DVQVAKKLAKLKNKKVVVVKTKEGLVPALTSKGLDMIAGMSPTERRKKEINFSKPYYI 123
 DV + + + + V + ++G++PAL S + I + + T ER + ++PS PY+
 Sbct: 69 DVDLIQAIGEAQVITVDIQGYPDFGIIIPALQSNIVGAISAITITPERAQSVSPSPYFK 128

Query: 124 SEPTLVVNABGKYTNAKNISDFKNKAVTAQGVLYLNLIDQINGVKKEVAMGDFNQLRQA 183
 S L + + + KN+ D + ++ G + + G K + +F+ + A
 Sbct: 129 S--VLAIAVQDGNDTIKNLKDEGKRLAVAIGTIGAMVATNVPGAK---VTFNDSITSA 182

Query: 164 VE---SGVWDAYVSEKRPDATSAQTANPKLWMIELHQGFKTSDADTNISVGMKGNKRNQ 240
 ++ +G DA ++RP A + L+ ++ + D I++ + INQ
 Sbct: 183 LQELVNGADAVINDRPFVLLYA-ICKDGLNVKLSADVGSEDY-YGIAMPLAP-PGRINQ 239

Query: 241 VNQVL----ESISRDKQIALMDKMKIEQ-----PSV-----KKEKNGK 274
 +VL + I A+ +K E+ PS+ + N
 Sbct: 240 TREVLNQLFLQIENGTYNAIYEKWFGEKKNPFLPLVAPSLVGKVTAQSITPERSQNEN 299

Query: 275 PNFFEQMATILKNNGSQFIRGTATLLISNVGTIVGLFGLLIGVFRTAPKSDNKLKAL 334
 NF + T+ +N +G+ T+L++ GL G + + A SD
 Sbct: 300 DNF---LITLERN---LFXGSILTVLLTAFSVFFGLIGGTGVAI---ALISD----- 342

Query: 335 QKLLGWLNLNIEVIEFPGTTPMIVQSMVIYYOTAQAF-----GVSLDRTLAAIFIVSINTGA 389
 K L + IY+E FRGTPM+VQ +IY+G F G++DR AAI +S+N A
 Sbct: 343 IKPLQLLIRIYVEFRGTPMLVOLFLIYFGLPALFKIEGLGITIDRPPAAIALSLSLVA 402

Query: 390 YMSEIVRGGISFVDKQGFEAATALGSMTHGQTMKRVLPQVVRNIPATGNEFVINIKDTS 449
 Y++EI+RGGI S+D+GQ+EA +LG + QTM++++ PQ R ILP GNEF+ IKDTS
 Sbct: 403 YLARIIRSGIQSIDQGNQNEACSLGMSFWQTMKRVLPQAFRRILPPLGNEFTLLIKDTS 462

Query: 450 VLVNIVSVVELYPSGNTVATQTYQYFQTFTIIAILIYFILFTVTRILRYIEKRED 503
 + VI EL+ G + TY+ F+ +A++Y +LT + ++E D
 Sbct: 463 LTVAVIGQELFREGQLIVATTYRAFEVYIAVALVYLLLTTSISFVFKWLENYMD 516

There is also homology to SEQ ID 1194.

A related GBS gene <SEQ ID 8923> and protein <SEQ ID 8924> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 2
 McG: Discrim Score: 6.23
 GVH: Signal Score (-7.5): 0.11
 Possible site: 24

>>> seems to have a cleavable N-term signal seq.
 ALOM program count: 3 value: -8.86 threshold: 0.0
 INTEGRAL Likelihood = -8.86 Transmembrane 301 - 317 (295 - 324)
 INTEGRAL Likelihood = -6.05 Transmembrane 479 - 495 (473 - 496)
 PERIPHERAL Likelihood = 1.32 441
 modified ALOM score: 2.27

*** Reasoning Step: 3

----- Final Results -----
 bacterial membrane --- Certainty=0.4545 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

34.3/57.3% over 462aa

Synechocystis PCC6803

EGAD|48193| glutamine-binding periplasmic protein/glutamine transport system permease protein Insert characterized

-2111-

GP|1652664|dbj|BAA17584.1|D90907 glutamine-binding periplasmic protein [Synechocystis
sp.] Insert characterized
PIR|S77250|S77250 hypothetical protein - Synechocystis sp. (strain PCC 6803) Insert
characterized

EF071242 (454 - 1809 of 2146)
EGAD|46193|all1270 (54 - 516 of 530) glutamine-binding periplasmic protein/glutamine
transport system permease protein [Synchocystis PCC6803]GP|1562564|dbj|BAA17564.1||D90907
glutamine-binding periplasmic protein [Synchocystis sp.]PIR|S77250|S77250 hypothetical
protein - Synchocystis sp. (strain PCC 6803)
%Match = 12.3
%Identity = 34.2 %Similarity = 57.2
Matches = 128 Mismatches = 149 Conservative Subs.s = 86

204 234 264 294 324 354 384 414
PSEVCIPE*HKNTINRFO*NDIRIDLVFR*NRRK*LIGGC*MKKILLSLFTALLITXGOMTSIOADEVLEVRGMEAAAP

MKGIVKLGHWGKTWRYLLALGVLLAIAIPLLPAFSQVS
10 20 30 40

[illegible][illegible]

903 957 987
 RP-----DATSAQTANPKLK-MIELHQG-FKTSADATNLSV
 || || || : : : ||
 RPVLLYAIKDGALRNKISADV----NPPFLFLVAPSLVGKVGTAQSLTSPQANPNDFLLTFLRNLFKGS-----
 210 270 280 290 300 310

```

1017      1047      1077      1107      1137      1167      1197      1227
GMRKGDNRINQVQVLESIRDKQIALMDKMIKBQPSVKKEKNGKNPFEBQMATILKNNGSQFLRGTTATLLISMVGTIV
      : : : : |
-----ILTVLLTAF
                        320

```

[illegible][illegible][illegible]

There is also homology to SEQ ID 5804.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2112-

Example 1867

A DNA sequence (GBSx1974) was identified in *S. agalactiae* <SEQ ID 5807> which encodes the amino acid sequence <SEQ ID 5808>. This protein is predicted to be ATP-binding. Analysis of this protein sequence reveals the following:

5 Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.3208 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:CAE73160 GB:AL139076 putative glutamine transport ATP-binding
protein [Campylobacter jejuni]
Identities = 132/241 (54%), Positives = 178/241 (73%), Gaps = 1/241 (0%)

Query: 5 ILEIKHLKSYSGSNEVLKDISLVNKGVEISITGSSGSGKSTFLRSINLLEEPSGGEILY 64
++E+K+L+K YG EVLAK+I+ ++KG+VI+IIG SG GKSTFLR IN LE GEIL
20 Sbjct: 1 MIEVNLQKKYGEVLNINNTISKGDVIAIGPSGGGKSTFLRCINRLKLEADSGEILY 60

Query: 65 HGRNVLKKGVDLNNYREKLGVMFQSFNLFENLNILENAIVAQTIVLKRERQEAELAKEN 124
+ N+L+K D+N R+K+ MVFO FHLF N N+EN + ++EA K AK
25 Sbjct: 61 NKQNLDKKEDINKIRQKVMFQHFNLFANKNVNMEMLCLATPIKGTILSGEAKKAKLL 120

Query: 125 LNAVGMTEQYWKAKPQLSGGQKQKVAIALARALSVNPEAILDFEPTSLALDPVMVEVLKTM 184
L VG+ ++ P +LSGGQKQK+ALAR+L +NP+ ILDFEPTSLALDPVM+GEVL M
30 Sbjct: 121 LAKVGLADKE-NIMPHKLSGGQKQKVAIALARALSVNMPDVLDFEPTSLALDPVMIGEVL SIM 179

Query: 185 QDLAKSGLTMIIVTHEMEFAKVSDDRVIIFMDKGIIABQGTTFQGLFNPNTQERTKEFLQRL 245
+D+AK GLTM+VTHEM FA+ V+R+ FMDKG IA +FK+FNFP+ ER +EFL + L
35 Sbjct: 180 KDVAKEGLTMLVTHEMGFARKVANRIFFMDKKGIAVDASPKVEFNPNSNERLREELKVL 240

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2157> which encodes the amino acid
35 sequence <SEQ ID 2158>. Analysis of this protein sequence reveals the following:

Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
40 bacterial cytoplasm --- Certainty=0.1170 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

45 Identities = 212/246 (86%), Positives = 237/246 (96%)

Query: 1 MTQAILEIKHLKSYSGSNEVLKDISLVNKGVEISITGSSGSGKSTFLRSINLLEEPSGG 60
M+ ++I+EK+LKKSYSGSNEVLKDISLVNKGVEISITGSSGSGKST LRSINLLEEPS G
50 Sbjct: 24 MNSIIEIKHLKSYSGSNEVLKDISLVNKGVEISITGSSGSGKSTFLRSINLLEEPSAG 83

Query: 61 EILYHGRNVLKKGVDLNNYREKLGVMFQSFNLFENLNILENAIVAQTIVLKRERQEAELI 120
+IL+HG +VL + Y+L +YREKLGVMFQSFNLFENLN+LENAIVAQTIVLKR+R +AE+I
55 Sbjct: 84 QILFEGEDVLAKEHYNLTHYREKLGVMFQSFNLFENLNILENAIVAQTIVLKRDRQAQEQI 143

Query: 121 AKENLNAVGMTEQYWKAKPQLSGGQKQKVAIALARALSVNPEAILDFEPTSLALDPVMVEV 180
AKENLNAVGMTEQYWKAKPQLSGGQKQKVAIALARALSVNPEALDFEPTSLALDPVMVEV
60 Sbjct: 144 AKENLNAVGMTEQYWKAKPQLSGGQKQKVAIALARALSVNPEALDFEPTSLALDPVMVEV 203

Query: 181 LKTMQDLAKSGLTMIIVTHEMEFAKVSDDRVIIFMDKGIIABQGTTFQGLFNPNTQERTKEF 240
LKTMQDLAKSGLTMIIVTHEMEFA++VSDR+IFMDKG+I E+G+P+Q+FNPTQ+RTKEF
Sbjct: 204 LKTMQDLAKSGLTMIIVTHEMEFARDVSDRIIFMDKGLITRESGPQIFNPNTQDRITKEF 263

-2113-

Query: 241 LQRFLK 246
 LQRFLK
 Sbjct: 264 LQRFLK 269

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1868

A DNA sequence (GBSx1976) was identified in *S. agalactiae* <SEQ ID 5809> which encodes the amino acid sequence <SEQ ID 5810>. This protein is predicted to be hypersensitive-induced response protein. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -17.94 Transmembrane 4 - 20 (1 - 28)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.8175 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9479> which encodes amino acid sequence <SEQ ID 9480> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GF:AAF68390 GB:AF236374 hypersensitive-induced response protein
 [Zea mays]
 Identities = 127/275 (46%), Positives = 174/275 (63%), Gaps = 1/275 (0%)
 Query: 19 ITSLLYVVEQCTVAI IERFGYKQTATSGIHIVPLGIDKIAARVQLRLQSBIIIVETRTK 78
 I L V Q TVAI E FGK + G H +IA + LR+ Q ++ ETRTK
 Sbjct: 4 ILGLVQVDQSTVAIKENPGKFSEVLEPGCHFLPWCIGQQIAGYLSLRVQLDVRCTRTK 63
 Query: 79 DNVFVTLNIATQYRVNERNVTDAYYKLIKPEAQIKSYIEDALRSSVPKLTDELFEKDE 138
 DNVFVT + QYR + +DA+YKL QI+ST+ D +R++VPEKL LD+ FE+K+E
 Sbjct: 64 DNVFVTVASVQYRALADKASDAFYKLSNTRQIQSTVFVDVIRATVFKGLDDAFPEKNS 123
 Query: 139 IALEVQHQAEEEMSTGYIIVKTLITKVEPDAEVKQSMNEINAAQRVRAAQLANADKI 198
 IA V+ ++ + MSTGY IV+TLI +EPD VE++MNEINAA R RVAA E A+KI
 Sbjct: 124 IAKAVEEELEKAMSTGYQIVQTLIVDISEPDRVGRAMNEINAAARMVRAASEKAEKI 183
 Query: 199 KIVTAAEAEDAKDRHLHGVGIAQQRKAIVDGLADSIQLKIDANVLTLEQIMSLITNQYL 258
 + AE EAE L GVGIA+QR+AIVDGL DS+ + T + IM ++L QY
 Sbjct: 184 LQIKKAGEAESXYLAGVGIARQRQAIVDGLRDSVLAFSENVPGITAKDIMDLVLTQYF 243
 Query: 259 DTINIF-AINGNQITFLPNPBGVEDIRTVLSAL 292
 DT+ A + + ++P+P P V+D+ Q+ L
 Sbjct: 244 DTWREIGASSKSSSVFIHPGPGAVKQVSAQIRIDGL 278

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 5811> which encodes the amino acid sequence <SEQ ID 5812>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -13.06 Transmembrane 5 - 21 (1 - 29)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.6222 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

-2114-

The protein has homology with the following sequences in the databases:

>GP:AAF68390 GB:AF236374 hypersensitive-induced response protein
[Zea mays]
Identities = 126/273 (46%), Positives = 174/273 (63%), Gaps = 3/273 (1%)

Query: 23 LYVVRQQSVAIVERFGYRQKTATSGIHRLPFGI-DKIAARVQLRLQSEIIVETKTKDN 81
L V Q +VAI E PG++ G H LP+ I +IA + LR+ Q ++ ETKTKDN
Sbjct: 7 LVQVDGSTVAIKENPGKFSVLREPGCHF-LPWCIQQQIAGYLSLRVQLDVRCEETKTKDN 65

Query: 82 VFVTLNVAQYRVNEQNVDAYYKMKPESQIKSYIEDALRSSVPKILTDELFEKKDBEIA 141
VFVT+ + QYR +DA+YKL QI+SY+ D +R++VPEL LD+ FE+K+EIA
Sbjct: 66 VFVTVVASVCYRALADKASDAFYKLSNTRBQIQSYVFDVIRATVPKLGLDADFQKNEIA 125

Query: 142 LEVQHQVAEMSTYGYIIVKTLITKVEPDAVRQSMNEINAAQRKRVAAQELANADKIKI 201
V+ ++ + MSTYGY IV+TLI +EPD VK++MNEINAA R KVAA E A A+KI
Sbjct: 126 KAVEEBLEKAMSTYGYIQVQTLIVDIEPDDRVKRAMNEINAAARMVAASEKAEAKILQ 185

Query: 202 VTAAABAERKDRLHGVGIAQQRKAIVDGLAESIQELKEANISLNEBQIMSILLTNQYLD 261
+ AE EAE L GVGI+QR+ATVDGL +S+ E + IM ++L QY DT
Sbjct: 186 IKQAGBGAESKYLAVGVIARQCAIVDGLRDSVLAFSENVPQTAKIDMNVLTQYPTD 245

Query: 262 INTFAAKG-NQTLFLPNTPSGVEDIRTVLSAL 293
+ A + ++F+P P V D+ Q+ L
Sbjct: 246 MRBTGASSKSSSVFIPIHGFQAVKDVSAQIRGL 278

An alignment of the GAS and GBS proteins is shown below.

Identities = 254/291 (87%), Positives = 278/291 (95%)

Query: 5 IILTVILLVILVLLITSLYVVKQQTVAIIRFGYRQKTATSGIHRLPFGIDKIAARVQL 64
I + +++++ ++ ++LYVV+QQ+VAI+ERFG+YQKTATSGIHIR+P GIDKIAARVQL
Sbjct: 6 IPIAFGVIVILAVASTLLYVVKQSVAIIRFGYRQKTATSGIHRLPFGIDKIAARVQL 65

Query: 65 RLQSEIIVETKTKDNVFPVILNIATQYRVNENVDAYYKIKPEAQIKSYIEDALRSSV 124
RLQSEIIVETKTKDNVFPVILN+ATQYRVNE NVTDAYYK+KPE+QIKSYIEDALRSSV
Sbjct: 66 RLQSEIIVETKTKDNVFPVILNVATQYRVNEQNVDAYYKMKPESQIKSYIEDALRSSV 125

Query: 125 PKLTLDELFEKKDBEIALEVQHQVAEMSTYGYIIVKTLITKVEPDAVRQSMNEINAAQR 184
PKLTLDELFEKKDBEIALEVQHQVAEMSTYGYIIVKTLITKVEPDAVRQSMNEINAAQR
Sbjct: 126 PKLTLDELFEKKDBEIALEVQHQVAEMSTYGYIIVKTLITKVEPDAVRQSMNEINAAQR 185

Query: 185 KRVAAQELANADKIKIVTAAEABAERKDRLHGVGIAQQRKAIVDGLADSIQELKDAVILT 244
KRVAAQELANADKIKIVTAAEABAERKDRLHGVGIAQQRKAIVDGLA+SIQELK+AN++L
Sbjct: 186 KRVAAQELANADKIKIVTAAEABAERKDRLHGVGIAQQRKAIVDGLAESIQELKEANISLN 245

Query: 245 EBQIMSILLTNQYLDITNTFAINGNQTIFLNPPEGVEDIRTVLSALKTR 295
EBQIMSILLTNQYLDITNTFA QNOT+FLPN P GVEDIRTVLSALKTR+
Sbjct: 246 EBQIMSILLTNQYLDITNTFAAKGNQTLFLNPPEGVEDIRTVLSALKTK 296

SEQ ID 5810 (GBS231) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 7; MW 60.9kDa).

GBS231d was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 155 (lane 5-7; MW 59kDa) and in Figure 239 (lane 11; MW 59kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 155 (lane 9; MW 34kDa) and in Figure 183 (lane 6; MW 34kDa). Purified GBS231d-GST is shown in Figure 246, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2115-

Example 1869

A DNA sequence (GBSx1977) was identified in *S. agalactiae* <SEQ ID 5813> which encodes the amino acid sequence <SEQ ID 5814>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 34
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2305 (Affirmative) < succ>
10     bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9291> which encodes amino acid sequence <SEQ ID 9292> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

15     >GP:CAB13457 GB:Z99112 similar to hypothetical proteins [Bacillus subtilis]
      Identities = 259/514 (50%), Positives = 350/514 (67%), Gaps = 9/514 (1%)

      Query: 1  MGMTMENGAKKESDKPATTVGEVGQILSKGVLMGARGNSGVITSLQFRGPGQSIKKEEL 60
      M ++M +GA+EV +G+VG LSKG+LMGARGNSGVITSLQFRGPF ++I+ K+E+
20     Sbjct: 46  MNLSTMSGAREVEYQMTDDIGKVGSLSKGLLMGARGNSGVITSLQFRGPFKNITETKKEI 105

      Query: 61  TGQDLAHAFQNGVEVAYKAVMKPVEGTILTIVSRGAATAALKKAEITDDAVEVMRATLKGA 120
      + A A Q GV++AYKAVMKPVEGTILTIV++ AA A+ AB+ D ++M A + A
25     Sbjct: 106  NALFPAALQAGVDMAKAVMKPVEGTILTIVAKDAAKKAMILAEITDITAMTAVTEEA 165

      Query: 121  KRALAKTFMFLPVLKGVGVVDSGGQGLVFIYBGFSLTGEYIASEDFKATPAIMTEVNH 180
      + +L +TP++LFLKGVGVVDSGG+GL+ +YBGF++L GE + KA ++ +MV+
30     Sbjct: 166  EASLAKTIFELLFVLKGVGVVDSGGGLLCVYBGFSLAKGETVFPQ---KAVPLSDHDHS 222

      Query: 181  AEHHKAVNVGHVATEDIKYGYCTEVMGKLGKQGYTYKEFNYEERPGYLSNGLDSLLVAND 240
      AEHHK+ + TEDI++G+CTEVMV L Q +EP+ F+ LS GDSLLN+ D+
35     Sbjct: 223  AEHHKSAQMMNTSDIEFGFCTEVMVRLDQTK---REFDSTFRQDLQSGGDSLLVADE 279

      Query: 241  EIVKVFVITEDPGLVMGSLKYGSLKVKVKNMNMNHDA---QMKEVESETVKETKEYG 297
      + KVF+H E+PG V+ YG L+K+K+ENNR QH + Q K ET + YG
40     Sbjct: 280  SLAKVHIHASEPGNVLYAQHYGELIKIKIENMRQHTSIISQESKPADNETPPAKQPYG 339

      Query: 298  IIAVVGADGLAEIKFSQGVVDYIISGGQTMNPSTEDIVKAEKVARNRVILPANNHIA 357
      I+ V G+G+A++FKS G +I GQQTMNPSTEDIV A++ VNA V ILPNN NI NA
45     Sbjct: 340  IIVTVAMGEIADLFKSGISGVVIEGGQTMNPSTEDIVDAKSVNADTVFIPLNNSNIHA 399

      Query: 358  AQAADVVDIPAAVVEITKVPQGSPTSLIARDEPAKSLKETNVADMTNLSLDSVIGSVPLAVR 417
      A AA VVD V+ +TVPGG ++LLAF+P + E N A+M +++ V SG VT +VR
50     Sbjct: 400  ANQAASVVDQVFPVPIAKTVPQMSALLAFNPQDAEAMNANMLGAIQQVKSQGVTFVSR 459

      Query: 418  DTTIDGLETHENDILGMDGKILVSTPDMKALKDTFDKMDIEDSEIVTVYVKGEGKQAL 477
      DT IDG +I + D +G++G I+ ++ + A K +MI ED EIVTY CED Q
55     Sbjct: 460  DTHIDGDIKKGDFMSILNGTTIOTSENQLSAAKMLLSMTGHDDEIVTVYLYGHDASQEE 519

      Query: 478  AETLSVELEETVEDEVEIHQGDQPVVYILMSVE 511
      AE L +L E YE+EVRIH G QP+Y Y+SE E
60     Sbjct: 520  ABQLEAPLSEKYEKIRVKIHNGKQPIYSYVSAR 553

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 5635> which encodes the amino acid sequence <SEQ ID 5636>. Analysis of this protein sequence reveals the following:

```

      possible site: 15
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1816 (Affirmative) < succ>
60     bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

```

-2116-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 434/511 (84%), Positives = 475/511 (92%)

5 Query: 1 MMTMNGAKGVSDKPAITVGEVQQLSKGVLMGARGNSGVITSQLFRGFGQS IKDKDEL 60
M MMTMNGAKEV+DKPA+TVGEVQQLSKGVLMGARGNSGVITSQLFRGFGQS IK K+EL
Sbjct: 44 MMTMNGAKEVADKPAITVGEVQQLSKGVLMGARGNSGVITSQLFRGFGQS IKDKDEL 103

10 Query: 61 TGGDLAHAFQNGVEVAYKAVMKPFVEITILTVSRGAATAALKUGAETDDAVEVMRAITLKG 120
TG+DLA AFQ GVEVAYKAVMKPFVEITILTVSRGAATAALKUGA+ TDDAVEVM+A L GA
Sbjct: 104 TGGDLAQAFCQNGVEVAYKAVMKPFVEITILTVSRGAATAALKUGADLTDDAVEVMAALDGA 163

15 Query: 121 KRALAKTPDMLPVLKEVGVDSGGQSLVFIYEGFLSALTGEYLIASEDFKATPATMTEMVN 180
K ALAKTPD+LPVLKEVGVDSGGQSLVFIYEGFLSAL G+Y+ S DFKATEPA M+EM+N
Sbjct: 164 KRALAKTPDILLPVLKEVGVDSGGQSLVFIYEGFLSALMGDYYTSADFKATPANMSEIN 223

20 Query: 181 AEHHKAVVGHVATEDIKYGYCTEVMVGLKQGFTYVKEFNYEEFGYLSNLGDSILLVNDD 240
AEHHK+VVGHVATEDI YGYCTE+MV LKQGFTYVKEFNY+EEFGYLS LGDSILLVNDD
Sbjct: 224 AEHHKSVVGHVATEDITYGYCTEIMVGLKQGFTYVKEFNYEEFGYLSNLGDSILLVNDD 283

25 Query: 241 EIVKVRVHTEDPGLVMQGLKYGSLVKVKNMRNQHDAQMKVEVEETVKETKEYGIIA 300
EIVKVRVHTEDPGLVMQGLKYGSL+K+KV+NMNRNQH+AQ+CK +VE+ E K+G+IA
Sbjct: 284 EIVKVRVHTEDPGLVMQGLKYGSLIKIKVNMNRNQHQAQCKTDDVEKNKAVIDGLGIIA 343

30 Query: 301 VVAGGLAEIFKSGVDYIISGGQTMNPSTEDIVKAIKRNARNVILPNNKNI FPAAGS 360
VVAG+GL+EIFK+QGVVDY+ISGGQTMNPSTEDIVKAIK VNA+ VILPNNKNI FPAAGS
Sbjct: 344 VVAGGLAEIFKAGVDYIISGGQTMNPSTEDIVKAIKRAVNAQVILPNNKNI FPAAGS 403

35 Query: 361 AADVVDIPAAVVEITRVFGFTSLLAFLDPAKSLTNVADMTNLSLSDVSGSVTLAVRDTT 420
AA+VVDIPAAVV TRTVFGFTSLLAFLDPAKSL NVDN+ SLSDV+SGSVTLAVRDTT
Sbjct: 404 AADVVDIPAAVATRVFGFTSLLAFLDPAKSLTNVADMTNLSLSDVSGSVTLAVRDTT 463

40 Query: 421 IDGLEIHHNDILGMVDGKILVSTPDMKALKDTDFEMIDEDSEIVTIYVGEDGKQALAE 480
IDGLEIHHND LGMVDGKI+VS PME LK F+EMIDEDSEIVTI+VGE+G Q LAE
Sbjct: 464 IDGLEIHHNDFLGMVDGKIIVSNFDMKATLKAAFEMIDEDSEIVTIYVREGDGDLAE 523

Query: 481 LSEYLEETVEDVEVEIHQGDQPVYPYLMVSE 511
L+ YL ETYEDVEVEIHQGDQPVYPYLMVSE
Sbjct: 524 LAGYLGETYEDVEVEIHQGDQPVYPYLMVSE 554

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1870

45 A DNA sequence (GBSx1978) was identified in *S. agalactiae* <SEQ ID 5815> which encodes the amino acid sequence <SEQ ID 5816>. Analysis of this protein sequence reveals the following:

Possible site: 46
>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4771 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

55 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2117-

Example 1871

A DNA sequence (GBSx1979) was identified in *S. agalactiae* <SEQ ID 5817> which encodes the amino acid sequence <SEQ ID 5818>. This protein is predicted to be proliferating-cell nucleolar antigen P120. Analysis of this protein sequence reveals the following:

```

5   Possible site: 55
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10  bacterial cytoplasm --- Certainty=0.3774 (Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9345> which encodes amino acid sequence <SEQ ID 9346> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC74905 GB:AB000278 putative nucleolar proteins [Escherichia
   coli K12]
   Identities = 87/229 (37%), Positives = 128/229 (54%), Gaps = 8/229 (3%)

20  Query: 63  GKSIEHTGLVYSQEPAAQ--IVAQIAEPQSGMKVLDLAAAPGGKTHLLSYLNTGLLV 120
   G + EH +GL Y QE ++ + A A+ +V+D+AAAPG KTT ++ +NN G ++
   Sbjct: 89  GSTAEHLGLFYIQEASMLPVAALPACNAPQRMVDAAPGSKTTQISARMNIGAIL 148

   Query: 121 SNEISNKRSKILVNVERFGARNVIVINESQRKACFNSFFDLIVFDGPGSGEGMFRKD 180
   +NE S R K+L N+ R G NV +T+ + FD I+ D PCSGEG+ RKD
   Sbjct: 149 ANEFSASRVKLVHANISRCGINSVALTHFDGRVFGAAVPEMPDAILLDAPCSGEGVVRKD 208

   Query: 181 PQAIQVHKDYPTECAQLQDILKRAIKMLAHGGILVYSTCTVSPEEREEVNNQLQBY- 239
   P A++ W + E A QR++ A L GG LVYSTCT + ENE V WL + Y
   Sbjct: 209 PDALGNWSPESNOEIATQRELDISAFHALRPGGTLVYSTCTLNQRENEAVCLMLKETYP 268

   Query: 240 ---DYLELVDIPKLNQWVEGINVPQVARMYPHHFGQGGQFVAKLEDTRS 285
   ++L L D+ G + + ++P + SG FVA+LR T++
   Sbjct: 269 DAVEFLPLGDL--FPGANKALTEGFLHVFPPQIYDCGFFVARLKTQA 315

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 5819> which encodes the amino acid sequence <SEQ ID 5820>. Analysis of this protein sequence reveals the following:

```

Possible site: 34
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
40  bacterial cytoplasm --- Certainty=0.2316 (Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 213/311 (68%), Positives = 254/311 (81%), Gaps = 3/311 (0%)

50  Query: 1  MKLPNEFTIEKYQITLKDEAEAFDSEQKPIISAYRTNLEKKGOLDPNAIPSTPWGHYKG 60
   M LP EFI YQ IL E E P SF Q+P++A+R NPLK + F + IP+T WG+YKG
   Sbjct: 2  MSLPKPEINTYQAILKGELEDFIASFNGRPFVNAFRINFLKNQKLTFFHPDPNTLNGYHKG 61

   Query: 61  ISGKSIEHTGLVYSQEPAAQIVAQIAEPQSGMKVLDLAAAPGGKTHLLSYLNTGLLV 120
   +SGKS EH +GLVYSQEPAAQ+VAQ+A PQ+G +VLDLAAAPGGK+THLL+YL+NTGLLV
   Sbjct: 62  LSGKSPHVSGLVYSQEPAAQMVAVQAPQKGSRLDLAAAPGGKSTHLLAYLNTGLLV 121

   Query: 121 SNEISNKRSKILVNVERFGARNVIVINESQRKACFNSFFDLIVFDGPGSGEGMFRKD 180
   SNEIS KRK+LVEN+RRFGARNV+VINES+ RLAK F+ +FD IVPDPCSGEGMFRKD
   Sbjct: 122 SNEISKRSKVLVNIERFGARNVVINESADRLAKVFSHPYDITVPDPCSGEGMFRKD 181

```

-2118-

Query: 181 PQAIQYWHKDYPTKCAQLQRDILKKAIKMLAHGGILVYSTCTWSPEENEEVVMILLQEYD 240
 P AIQYWH YP ECA+LQ+ IL+AA+ ML GG L+YSTCTW+DEENK+VV WLL+ Y
 Sbjct: 182 PDAIQYWHHGYPAECALQKSLKEDALAMLPKGELIYSTCTWAPENEDVVMILLBTYT 241

Query: 241 YLELVVDIPKLNQMGVGINVPOVARMYPHHPQEGGFVAKLRDTRSKQAQIKPKPAQKIN- 299
 +LELVVD+PKLNQMGV GI +P+ ARMYPH +QEGGFVAKL+D R +E Q K KA K N
 Sbjct: 242 FLELVVDVPKLNQMGVSGIGLPETARMYPHYQEGGFVAKLRDKR-QEGGSTKLKAPKNSL 300

Query: 300 -KMQQLQWQOF 309
 K QL+LW+ F
 Sbjct: 301 IKDQLRLWKMP 311

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1872

A DNA sequence (GBSx1980) was identified in *S.agalactiae* <SEQ ID 5821> which encodes the amino acid sequence <SEQ ID 5822>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4111(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC24940 GB:AF012285 unknown [Bacillus subtilis]
 Identities = 86/240 (35%), Positives = 133/240 (54%), Gaps = 10/240 (4%)

Query: 6 DPAKQLVYKAGQFIKSEMQNTFDVEEKSFFDDLVTSLDKKTKLLIQEIIQHPYDNLILA 65
 + AK+ + AG I M + +E KS +DLVT+DK+T+K I I + +P IL
 Sbjct: 9 EIAKQWIREAGARITQSMHESLTITKSNFNDLVNIDKETKFPFIDRIQETPFGRILIG 68

Query: 66 EE---DEVRSPIAQGNVWVLDPIDGTVPFVQKINFVAMLAYEEGVGQFGIYDVNADI 122
 EE D + S +G VW++DFIDGT+NF+ Q+ NFA+ + +E G G+ G+IYDV+ D
 Sbjct: 69 EHQQDEKHS--LBGVWIIDPIDGTNPFVQQRNFAISIGIFENGEGKIGLIDVVDHDE 126

Query: 123 LYSGGGHFDVYANDKKIVPQRCPLERCLGVNSAMYAEN---DCGIAHLASLTGLVRI 178
 LY Y N+ K+ P +E +E +L N+ EN +A L G R
 Sbjct: 127 LYHAFSGRGAYNBTKLAPLAKETVIEELAINATVNIENRDRIDQSLAPLVKRVGRTS 186

Query: 179 YGQAGISMAKVMQGKLLAYPSY-IQPDVYAAKIMGETLGPILLTLDGSEPNYSTRQKVM 237
 YG A + +A V G++ AY + + PWDYAA ++ +G T T++GE + V+
 Sbjct: 187 YGSAALELANVAAGRIDAYITWRLAPWDVYAGCVLLNEVGVTYTTTIGEPPTPLNHSVL 246

A related GBS nucleic acid sequence <SEQ ID 10937> which encodes amino acid sequence <SEQ ID 10938> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5823> which encodes the amino acid sequence <SEQ ID 5824>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1843(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2119-

An alignment of the GAS and GBS proteins is shown below.

Identities = 155/253 (61%), Positives = 205/253 (80%)

```

5 Query: 1 NDAKFPDAKQLVYKAGQFIKSEMQNTFDVEKSRFDDLVTSLDKKIQKLLIQEIIQHYPD 60
++ K+ FA+Q++ +AG FIKS+M D++ K++FDDLVT++D++TQ+LL+ I Q YP
Sbjct: 8 LETKYAPARQIKKEAGLFIKSKMSQLDIQVKTCFDDLVTNVQDQEQQLLMDRHQYTPC 67

Query: 61 DNLIAEEDVRSPIAQGNVWLPIDGTNFIQKINFAVMLAYEESGVOQGLIYDVWA 120
D ILAE2++VR PI QGNVWV+DPIDGTNFIQ FAVM+AYYE+G+QQFG+IYDVWA
10 Sbjct: 66 DALLAEENDVRHPINQGNVWLPIDGTNFIQVCGSQFVMIAYYEQGLIYDVWA 127

Query: 121 DILYSGGSHFDVYANDKKIVPFCRCPLERCLLGVNSAMYARNDGIAHASETLGVRVY 180
D L+GGG F+V N K+ +QR PLER L+G N+ M+A ND +AHL ++TLGVR+Y
15 Sbjct: 128 DQLAGGGDFEVLINGDKLEAYQEKPLERSLIGCNAGMFARNDNRNLAHLIAKTLGVRVY 187

Query: 181 GAGISMAKVMQCKLLAYFSYIQWMDYAAAKIMGETLTPLTLTLDGEEPNYSTROKVMFLP 240
GAGI M KVM+ +LLAYFS+IQWMDYAAAK++G+ LG+ LIT+Dg EP+ TRQK+MF+P
15 Sbjct: 188 GAGICMVKMKQELLAYFSFIQWMDYAAAKVLGDKLGYLLTIDGYEFPDTRKIMFV 247

Query: 241 KSKLNLIQSFLT 253
K +L I S+LTK
20 Sbjct: 248 KQUTRIASFLT 260

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1873

A DNA sequence (GBSx1981) was identified in *S. agalactiae* <SEQ ID 5825> which encodes the amino acid sequence <SEQ ID 5826>. Analysis of this protein sequence reveals the following:

```

30 Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4131(Affirmative) < succ>
35 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC24938.GB:AF012285 unknwn [Bacillus subtilis]
Identities = 33/78 (42%), Positives = 50/78 (63%)

40 Query: 13 YSYPLDPSNNTEDI*KLRLFAQVHERAYENSIKVDLLDSYKRFKKVSKAQEKQIDRE 72
Y YP++ W TE+ V+ F QVE AYR ++LL +Y+ FK++V KA+EK++ R
Sbjct: 3 YQYPMNEDWTEAEVDVIAFPQQVELAYEKADREELLYKAYRRFRKIVPGKAEKKLOS 62

45 Query: 73 FORTSGYSTYQAVKAAQQ 90
F+ S YS Y+ VK A++
Sbjct: 63 FEEQSTYSYPTVKQARR 80

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 5827> which encodes the amino acid sequence <SEQ ID 5828>. Analysis of this protein sequence reveals the following:

```

Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
55 bacterial cytoplasm --- Certainty=0.4442(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

-2120-

Identities = 59/91 (64%), Positives = 70/91 (76%)

Query: 9 ISSNYSYFLDPSMNTEDITKVLRLNQLVHAYENSIKVDDLLDSYKPKPKVKGKAEKQ 68
 +S NY YPLD SW+TE+I+ VL FLN+VE AYE + LLD SYK +K +VKSQAQEKQ
 5 Sbjct: 5 MSGNYTYFLDLSMSTELISSVHLFLNKLVLAYEKKVDARQLLD SYKTYKTIVKSKAEKQ 64

Query: 69 IDREFORTSGYSTYQAVKAAQQQKGFISLG 99
 IDR+PQ+ SGYSTYQ VK A+ KGF SLG
 10 Sbjct: 65 IDRDFQKVSYSTYQVVKAKALIEGFFSLG 95

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1874

A DNA sequence (GBSx1982) was identified in *S.galactiae* <SEQ ID 5829> which encodes the amino acid sequence <SEQ ID 5830>. Analysis of this protein sequence reveals the following:

Possible site: 18
 >>> Seems to have no N-terminal signal sequence (or aa 1-18)

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0952 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF21893 GB:AF103794 unknown [Listeria monocytogenes]
 Identities = 74/126 (58%), Positives = 101/126 (79%)

Query: 1 MITLFLSPSCTSCRKAWLSKHEVAFEEHNIITSPLNKEELLQILSPTENGTEIISTR 60
 M+TL+ SPSCSTCRK+RAWL +H++ +E NI + PL+ +E+ +IL TE+GT++IISTR
 30 Sbjct: 1 MVTLYTSPSCTSCRKRAWLEHDIPIYKERNIPSEFLSLDEIKELRLMTEGTEIISTR 60

Query: 61 SKVPQKLAIDVDELSTSSIMELISENPFLRRPIILDKKRMQIGPNRDEIRAFLRDYRK 120
 SK PQKL +D+D L L ELI +NP LLRRPII+D+KR+Q+G+NEDEIR FLPR R
 35 Sbjct: 61 SKTFQKLNVDLDSFLQQLFELIQLNPGLLRRPIIIDEKRLQVGYNDEIRFLRRVET 120

Query: 121 QELKQA 126
 +L++A
 Sbjct: 121 YQLREA 126

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5831> which encodes the amino acid sequence <SEQ ID 5832>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0511 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 112/134 (83%), Positives = 127/134 (94%)

Query: 1 MITLFLSPSCTSCRKAWLSKHEVAFEEHNIITSPLNKEELLQILSPTENGTEIISTR 60
 M+TLFLSPSCTSCRKAWL KHEV F+EHNIITSP++EL+ ILSPTENGTEIISTR
 55 Sbjct: 1 MVTLYTSPSCTSCRKAWLVKHEVDQEHNIITSPSLRDEIMSLSPTENGTEIISTR 60

Query: 61 SKVPQKLAIDVDELSTSSIMELISENPFLRRPIILDKKRMQIGPNRDEIRAFLRDYRK 120
 SKVPQKL IDV+ELS S L++LI++NPFLRRPII+D+KRQIGPNRDEIRAF LR DYRK
 Sbjct: 61 SKVPQKLDIDVDELISDLDLIAKNPSLLRRPIIMQKRMQIGPNRDEIRAFLRDYRK 120

-2121-

Query: 121 QELQATIRAEIBG 134
 QEL+QATTI+AEIBG
 Sbjct: 121 QELQATIRAEIBG 134

SEQ ID 5830 (GBS232) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 10; MW 16.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 56 (lane 2; MW 42kDa).

GBS232-GST was purified as shown in Figure 207, lane 7.

- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1875

A DNA sequence (GBSx1983) was identified in *S.agalactiae* <SEQ ID 5833> which encodes the amino acid sequence <SEQ ID 5834>. Analysis of this protein sequence reveals the following:

- Possible site: 39
 >>> Seems to have an uncleavable N-term signal seq
- Final Results -----
- | | | |
|-------------------------|------------------------------|---------|
| bacterial membrane --- | Certainty=0.0000 (Not Clear) | < succ> |
| bacterial outside --- | Certainty=0.0000 (Not Clear) | < succ> |
| bacterial cytoplasm --- | Certainty=0.0000 (Not Clear) | < succ> |

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5835> which encodes the amino acid sequence <SEQ ID 5836>. Analysis of this protein sequence reveals the following:

- Possible site: 39
 >>> Seems to have no N-terminal signal sequence
- Final Results -----
- | | | |
|-------------------------|--------------------------------|---------|
| bacterial cytoplasm --- | Certainty=0.1768 (Affirmative) | < succ> |
| bacterial membrane --- | Certainty=0.0000 (Not Clear) | < succ> |
| bacterial outside --- | Certainty=0.0000 (Not Clear) | < succ> |

An alignment of the GAS and GBS proteins is shown below.

- Identities = 210/308 (68%), Positives = 252/308 (81%)
- Query: 1 MKIHINDYKDIQAKEDCVLVLGYFDGLHGHKALFDKAKKIATEKNLKIIVLTNETPR 60
 M+I YI DY+DI ++D VL+LGYFDGLH GHKALFDKA++A ++ LK+VV TF B+P+
 Sbjct: 1 MEIEYIKDYRINDQEDFTVLLIGYFDGLHGHKALFDKAREVANRGLKVVFVTFTEBPK 60
- Query: 61 LTPARFQPELLLHILTSPEKRSEKPKQRYGVDELILMNPTSHPSKVSSDLFIKKYTYGLRAK 120
 L F+RF PELLH+T F+KR EK+ +YGV++LYL++FTS PSKVSSD FI YI L+AK
 Sbjct: 61 LAFSRSSPELLLHITTPKRRYKFDYGVNKLVLVDFTSKPSKVSDDHETTHYIKMLAK 120
- Query: 121 AAVGFDYKFGHNRTSGDYLARNKGFVYIIDEISEGSEKISSTRIRQLITEGVVEKAKQ 180
 VVGFDYKFGHNRT DYL RN+G VY I+EI E KIS+T IR+LI RGNV KAN
 Sbjct: 121 HIVVGFDYKFGHNRTSDYLTNRNFSQVYIIEIKEDHRKISATWIRKLQEGNVVKANH 180
- Query: 181 LLYEFTSCGMVHEDARGRTIGFPTANLAPINRNTLPADGVYISNVLINGKYIRANTS 240
 LLSY+ ST G VHGEDARGRTIGFPTANLAPI+ TYLPADGVY+NV++ K YR+NTS+
 Sbjct: 181 LLSYDLSTRGVHGDARGRTIGFPTANLAPIDNTYLPADGVYVTVNVANKYIRANTS 240
- Query: 241 GKNITFGSTELRIEANTFDPDGDYQETIRIPWLKRIRSMVKVNGIDILVKQLKKDKETA 300
 GKN+TFGG ELRIE NIFDGD +IYQE IEI WL +IR+M KP GI+DL +L+ DK A
 Sbjct: 241 GKNVTFGGKELRIEIVNIFDDEEYIEIIEIVWLKIRDMKFEIGIEDLIRLEYDKRTA 300
- Query: 301 LANKKDSQ 308

-2122-

LNWKKDS+
 Sbjct: 301 LNWKKDSK 308

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1876

A DNA sequence (GBSx1984) was identified in *S.agalactiae* <SEQ ID 5837> which encodes the amino acid sequence <SEQ ID 5838>. This protein is predicted to be tRNA pseudouridine 5S synthase (truB). Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.2576 (Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 9817> which encodes amino acid sequence <SEQ ID 9818> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06129 GB:AP001515 tRNA pseudouridine 5S synthase [Bacillus halodurans]
 Identities = 145/283 (51%), Positives = 191/283 (67%), Gaps = 12/283 (4%)

Query: 2 ITGILNKKKNGMTSHDAVFKRLILTKKIGHGGTLDPOVGVLPVIAVGKATVIEVHT 61
 +TSI+ L K GMTSHD V KLR++L TTK+GH GTLDPOV GVLP+ +G AT+V +TH+
 Sbjct: 3 MTGILFLAKPRGMTSHDCVAKRLRLATKKVGHVGTLDPOVGVLPVFCIGHATKVAQYMS 62

Query: 62 ESKIVYEGSHITLGATSTHDSGGEVSKITPTLQSDLSQVVDHAKMSFTOPITQVPPMYS 121
 + K YEGE+T+G++T+TD SG+ + T Q E VVD + +F G I Q+PPMYS
 Sbjct: 63 DYPKAYEGEVTVGFSTTTDRSGDTVE-TKTIQQPFVEAVDQVLAFTVGEIKQIPPMYS 121

Query: 122 AVKNGHKLLEYARSGEEVERPKRQITISEFRRTSPLYFEKIGCRPSPVSCSKGTYVRT 181
 AVKV GK+LNEYAR+G VERP+R +TI R S + +E+G+CRF F VSCSKGTYVRT
 Sbjct: 122 AVKVRGRKLEYARAGITVERPERTVITPSLRMSDIVYEGVCRFPNVSCKGTYVRT 181

Query: 182 LAVDLGKIKGVASHMSFLKRTSSAGLSITQSLTLERINEKYKQ-EDPSFLPLPIEYGVLDL 240
 LAVD+G LGY +HMS L RT S S+ + T E+ E+ +Q E S LLPIE +LD+
 Sbjct: 182 LAVDIGKALGPAMMSDLVRTKSGPPSLEECFTPTLEERLEQYEGSSLLPIETAILDI 241

Query: 241 PKVNLTEEDKVEISYGR-----RIILENEADTLAAFYR 273
 P+V +E + +I +G R + NE L A Y+
 Sbjct: 242 PRVQVNEKEEKIKHGAVALPKQWPNHFRFTVYNEGALLAIYK 284

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5839> which encodes the amino acid sequence <SEQ ID 5840>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.2698 (Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 201/295 (68%), Positives = 246/295 (83%), Gaps = 2/295 (0%)

-2123-

Query: 1 MITGIINLKKEAGMTSHDAVFKLRKILFTKKIGHGGTILDPDVGVLPIAVGKATRVIEYM 60
 MI GIINLKKEAGMTSHDAVFKLRK+L KCIHGSGTILDPDVGVLPIAVGKATRVIEYM
 Sbjct: 1 MINGIINLKKEAGMTSHDAVFKLRKILQKKIGHGGTILDPDVGVLPIAVGKATRVIEYM 60

5 Query: 61 TESGKIYRGRTITLGYATSTEDS9GEVISRTPLTQSDLSEDVVDHAMKSFITGIPITQVPPMY 120
 TE+GK+YSG++TLGY+T+TED+SGEV++R+ L + L+B++VD M +F G ITQ PPMY
 Sbjct: 61 TRAGKVYBQQTILGYSTTTEDASGEVVARSSL-PAVITRELVDQ/TMTFTFLGKITQTPPMY 119

10 Query: 121 SAVKVNKKLYEYARSGERVERPKRQITISEFRRTSPLNYF-EKGICRFSFYVCSKGTIV 179
 SAVKVN+KLYEYAR+GE VERP+R++TIS F RTSPF F E G+CRPSF V+CSKGTIV
 Sbjct: 120 SAVKVNKKLYEYARAGESVERPRREVITISLFPRTSPINFTEDGLCRPSFKVACSKGTIV 179

15 Query: 180 RTLAVDLGKIKGYASHMSFLKRTSSAGLSITQSLTLBSEINRKYKQSDFSFLPIRYGVLD 239
 RTLAVDLG IG SHMSFL+R++SAGL++ + TL EI + +++ SFLPIRYGV D
 Sbjct: 180 RTLAVDLGRALGVESHMSFLORSASAGLTLETAYTLGRIADNVSKQMSFLPIRYGVAD 239

Query: 240 LPKVNLTEDKVEISYGRKILLNEADTLAIFYENRVAILEKRGNEPKPKVLL 294
 LPK+ + + + EIS+GR+ L ++ LAAP+ +VIALEKR E+KP KVL+
 Sbjct: 240 LPKVIDDTLTELISPGRLSLPSCQEPLLAAFEKGVIAILEKRDQEVKPKVLL 294

20

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1877

A DNA sequence (GBSx1985) was identified in *S. agalactiae* <SEQ ID 5841> which encodes the amino acid sequence <SEQ ID 5842>. Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2776 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9819> which encodes amino acid sequence <SEQ ID 9820> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12871 GB:299109 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 39/145 (26%), Positives = 68/145 (46%), Gaps = 7/145 (4%)

40 Query: 3 MKIRTATLDDSEKLVPLYQELG----YATLSERIQSILKVLTHSDYGLFIADNKGKLLA 58
 M I R A D+ + PL+ + A L ++ LK L + + LIAE+NG+ +
 Sbjct: 1 MNIRQAKTSDAAAIAPLNFQYREFYRQASDLQGAEPFKALENHESVILIAENGEFTIG 60

45 Query: 59 FVGYHKLYFFKSGITYRILALVNVNKHRRKGIAQLINHVKOLAKTDGSEVLAINSLK 113
 F + + + Y + L V R KG +L++ K A +G++L L + +
 Sbjct: 61 FTQLYPTFSVSMKRYIILNDLFVPIARTKAGGRLLSAADKYAGONGAKCLTLOT--E 113

Query: 119 EYRQEAHYFYNLGGKKVSTGFSYY 143
 + ++A YR G+++ TGF +Y
 Sbjct: 119 HHNRKARSLYONGYEE-DTGFVHY 142

50

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 5843> which encodes the amino acid sequence <SEQ ID 5844>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0962 (Affirmative) < succ>

-2124-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 37/126 (29%), Positives = 64/126 (50%), Gaps = 16/126 (12%)

Query: 18 PLYQE-----LGYAISLSBIQSILKVLTHSDYGFLLA--EDNKKLLAFVG--YHKLYF 67
P+ QE LGY +SL ++ + ++ + FL +D +LL +V Y LY
10 Sbjct: 11 PMLQBINAKLGYLVSLLDLERQVRLIEDCHYFTAYADKOTNQLGVYHARYETLY- 69

Query: 68 PEKSGTYRILALVNMKRRKGIASQLINHVQKLAKTGSEVIALNSSIKKRYQRAYHF 127
+ +L L V ++R+GI S L+ ++ A+ +G + LNS+ +R+EA+ F
Sbjct: 70 --ASDGINLLGLAVLPAYQRRGISALLRALESQARQBGIAFIRLNSA--SHRKAHAF 124

15 Query: 128 YENLGF 133
Y NL +
Sbjct: 125 YRNLDY 130

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
20 vaccines or diagnostics.

Example 1878

A DNA sequence (GBSx1986) was identified in *S.agalactiae* <SEQ ID 5845> which encodes the amino acid sequence <SEQ ID 5846>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1659 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
30 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

RGD motif 28-30

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:AAF30776 GB:AB002133 conserved hypothetical [Ureaplasma urealyticum]
Identities = 106/440 (24%), Positives = 206/440 (46%), Gaps = 65/440 (14%)

40 Query: 13 FAINESEYHQLLEQIRGNAPDFKEVSEERLEKERILIGEQANQLQEVVVE-KDKETALQY 71
F N+ +Y++L++Q +D LEK+R L E+ KN+ ++ KD + K
Sbjct: 71 FLANDRDYNELVKQ-----RYD-----LEKQDELKEKLNEGKNKAIAHPKDSPEYNKLI 120

Query: 72 KVKQFLIEKNLLKDNQYQLAECNLQKDMMLRD-----LENQIDRLRLHEHNSLQEA 123
K ++ + + + ME +++ ++ L+ L+N I + ++ +N+ + A
45 Sbjct: 121 KQSKINSLNKTIENESQYKKEIENIELKLSQPDERTSLENTAKQEIKLONAKQVA 180

Query: 124 IITKVERE-----RDAIQNLHLI-----RKEKDLALASVKSDY 156
+ + +D I + I+ E +K + + ++S
50 Sbjct: 181 IINFESNEVQKITKOKIDLDIEIKLKFAIQAHEDNMKAKENWSSKKIVEIKELBSKK 240

Query: 157 EVQLKAANEQVEFYKNFKQAQSKTAVGSELSHYASTFNKVRHLAFPNAYFEKDNLTSSR 216
++ +E +E K K+ + K VGE LE + + +F++ + P+ F K N
Sbjct: 241 DKEIHKLTESIBOLKREKSS-NVKLVGEELBONLKNKPDETYSFSCPDMTFTKINEAID- 298

55 Query: 217 GSKGDFPI-----REKDENDLEFL-SIMFEMQNESDITIKKHNEDPFKELDKDRREKS 269
G K DF+ +E +D + + S E K E D K KN +K+LD+DR +
Sbjct: 299 GKKAQDFLEFFDFGKSMENDDKLIGSATISAKTEFFDQKGTGNSAHYKKLDQKINQK 358

60 Query: 270 CEYAVLVIMLDEANDNYNTIGIVDSHKYPMYVIRPQFTLIGILGNAAALNTLYKQEL 329
EY+LVT LE ++ + + ++Y M+ +RPQ+FI L+ ++RN A TLK K
Sbjct: 359 SEYALLVTELEPEDHF---VKKICNEYKNMFAVRPQYPIPLVDMIRNFA--TLKAKINS 412

-1215-

Query: 330 ALMKEQNIDITHFEEDLDIFKNFAKYN-YNASKNFQKAIDRIDKSIKRMNAV-KAAL/T 387

+++ + D -EE+LD K N + +K ID+ IK+ E++ +AA

Sbjct: 413 QITRYE--DRAKIERMLDRLKCDIVDNTLKYINDTKKIIDSKATIKKASIEESASDI 470

Query: 388 SENQLRLANNKLLDDVSVKKL 407

+L K+++++K+

Sbjct: 471 INKKNTLKKKINELTIRKI 490

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5847> which encodes the amino acid sequence <SEQ ID 5848>. Analysis of this protein sequence reveals the following:

Possible site: 33

>>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3192 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 20 An alignment of the GAS and GBS proteins is shown below.

Identities = 310/445 (69%), Positives = 352/445 (78%), Gaps = 22/445 (4%)

Query: 1 MNEIKCFHCGTAFAINSEYHQLLEQIRGDAFDKEVSERLKRLILGSEQAKNQLQEVVV 60

MNEIKCFHC T F INSEY QLLDQ+RG APD+E+ +RL E +L E+AK+QL EVV

- 25 Sbjct: 1 MNEIKCFHCHTILFTINSEYSQLEQVRGQAFDESLKRLINETALLEEKAKHOLEVVA 60

Query: 61 KDKRIAKLQYKVKQF-----LTERKINL-----KINEYQLAEQLNQK 98

+K+ I L +++Q L +KD L+ N +LA QL +K

- 30 Sbjct: 61 KGETAITSLTINQLQTEKEQAYLRQELAKNDQLTASLEAKLDKLASQNALANQLAEK 120

Query: 99 DMMLRDLNQTDRLRLEHENSQLEALTKVERBRDAIQNQLHTQEKKDLALASVKSDEYV 158

D + L NQ+D+L LE + + Q L +E+ERD I+NQL +Q KE +L+LASV+SEYD

- 35 Sbjct: 121 DKEVVSILTQDKALEKDATPOSKLATIEKERDGIINQLAQAKSESLASVRSDEYA 180

Query: 159 QLKAAEQVEFYKNFKACQSTKA+GESLE YARTEFNKVR APNA F KDN LSSRGS 218

QLKAAEQVEFYKNFKACQSTKA+GESLE YARTEFNKVR APNA F KDN LSSRGS

- 35 Sbjct: 181 QLKAAEQVEFYKNFKACQSTKA+GESLE YARTEFNKVRAPNASFVKDNLSSRGS 240

Query: 219 KGDFTYREKDENDLSEFLSIFEMIONESDTHKKHNEDEPFKELDKDRREKSCYAVLVTH 278

KGD+TYRE D N +E LSIEMIONE+D T KHKN DFFKGLDKDRREK CETAFLV+M

- 40 Sbjct: 241 KGDFTYREVDANGVEILSIFEMIONEDATTHKHNSDFFKELDKDRREKCEYAVLVSM 300

Query: 279 LEADNDYNTGTVDVSHKYPKMYVIRPQFFIQLIGILRNALNTLYKQELALAIKEQNTD 338

LEADNDYNTGTVDVSH+Y KMYV+RPQ FTQLIGILRNALN+L YKQELAL+KEQNTD

- 45 Sbjct: 301 LEADNDYNTGTVDVSHKYPKMYVIRPQFTQLIGILRNALNSLHYKQELALVKQNTD 360

Query: 339 ITHFEEDLDIFKNFAKYNYSASKNFQKAIDRIDKSIKRMBAVKAALTSSENQLRLANNK 398

ITHFEEDLD IFKNFAKYN SAS NF+KAIDRIDKSIKRM VK LTTSENQLRLANNK

- 50 Sbjct: 361 ITHFEEDLDQFKNFAKYNQSASNFKKALIDRIDKSIKRMSEVVRPLTTSENQLRLANNK 420

Query: 399 LDDVSVKKLTRKNPTMKAKFDALD 423

L+DVSVKL+TR+NPTM+ KF+ALKD

- 50 Sbjct: 421 LEDVSVKLITRQNTMREKFEALD 445

- 55 SEQ ID 5846 (GBS304) was expressed in *E.coli* as a His-fusion product. The purified protein is shown in Figure 206, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2126-

Example 1879

A DNA sequence (GBSx1987) was identified in *S.agalactiae* <SEQ ID 5849> which encodes the amino acid sequence <SEQ ID 5850>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

```

5   Possible site: 34
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.1845 (Affirmative) < succ>
10  bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5851> which encodes the amino acid sequence <SEQ ID 5852>. Analysis of this protein sequence reveals the following:

```

15  Possible site: 51
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.2492 (Affirmative) < succ>
20  bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

25  Identities = 113/180 (62%), Positives = 141/180 (77%)

Query: 16  LSELVDCFKKAVPSKARAGDIRIINLSDMSPGLGIDYHNLRFFQDBQRSLKLYLQEGDV 75
      L  +VDCFKGKAV SK  GD+ +INLSDM LGI YH LRFTFQ ++R LL+YLL+GDV
30  Sbjct: 18  LGTVVDCFKGKAVSSKVVREVGVLINLSDMGTLGICVHQLRFTQCMRRQLRLYLEGDEV 77

Query: 76  LIASKGTVKKVAIFEEQDYPVVASANITILRPTQHIRGYLYLKFFDSEBQQQALENANKG 135
      LIASKGT+KKV +F +Q+ VVAS+NIT+LRP + +RGTY+K F DS GQ L+ A+ G
35  Sbjct: 78  LIASKOTLKVCVPHKQNRDVASSNITVLRPQKLEGTYYIKFFLDSPIGQALLDVADHG 137

Query: 136 KAVNNISTKELLNIAIPSIPLFRQDYLIQRYKQGLNDYRKRIARABQEWERIQNDRQQL 195
      K V+N+STKELL+I IP IPL +QDYLY Y +GL DY RK+ RABQWE IQN+I++ L
40  Sbjct: 138 KDVINLSTKELLDIPIPIPLWKQDYLINHYLRGLTDYHRKLNABQEWETIQNEIQKGL 197

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1880

A DNA sequence (GBSx1988) was identified in *S.agalactiae* <SEQ ID 5853> which encodes the amino acid sequence <SEQ ID 5854>. Analysis of this protein sequence reveals the following:

```

45  Possible site: 15
    >>> Seems to have a cleavable N-term signal seq.

    INTEGRAL  Likelihood = -7.43  Transmembrane  62 - 78 ( 55 - 82)
    INTEGRAL  Likelihood = -2.07  Transmembrane  130 - 146 ( 130 - 150)
    INTEGRAL  Likelihood = -1.28  Transmembrane  37 - 53 ( 37 - 53)

50  ----- Final Results -----
    bacterial membrane --- Certainty=0.3972 (Affirmative) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```


-2127-

A related GBS nucleic acid sequence <SEQ ID 9347> which encodes amino acid sequence <SEQ ID 9348> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5 >GP:CAA22372 GB:AL034446 putative transmembrane protein
  [Streptomyces coelicolor A3(2)]
  Identities = 38/139 (27%), Positives = 64/139 (45%), Gaps = 5/139 (3%)

Query: 15 SASVEILCRGWLIPVSATKYSKIVSVSISSIFGGLLSANNHVSLSIFNLCL-FGLFLS 73
      +A+ B++ RG L +      +++ ++ + FGL+H N +L + + G L+
10 Sbjet: 143 AATEVVPFGVLFRIIEEHIGTYIALGLTGLVFGIMHLNEDATLWGALAIIEAGFHLA 202

Query: 74 LTVILKGNINWGAOGINGAMNVCQGSVFGIIVSGEPMLSNLSLVHVKTYGADWISGKKGFE 133
      N+W G+H WN G VF VSG S L+ G ++GG FG E
15 Sbjet: 203 AAYATRNLM/LTIGVHFGWNFAGGVFSTVVSNGND-SEGILLDATMSGPKLLTGGDFGFE 261

Query: 134 GSMIT---SIVLIIVACYNL 149
      GS+ + ++L + WL
20 Sbjet: 262 GSVYSVGFGLVTLVFLWL 280

```

20 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1881

25 A DNA sequence (GBSx1989) was identified in *S.agalactiae* <SEQ ID 5855> which encodes the amino acid sequence <SEQ ID 5856>, which is a methylase gene homolog. Analysis of this protein sequence reveals the following:

```

Possible site: 33
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2192 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

35 RGD motif: 264-266

```

A related GBS nucleic acid sequence <SEQ ID 9929> which encodes amino acid sequence <SEQ ID 9930> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

40 >GP:BAA87672 GB:AB016260 Hypothetical gene, methylase gene homolog
  [Agrobacterium tumefaciens]
  Identities = 350/1230 (28%), Positives = 595/1238 (47%), Gaps = 99/1230 (7%)

45 Query: 1072 KEVARIEKGVDIRNAYQEVIAIQVYHYDKETPFNLHQLKLRITYDSFVKHYGLYLSAV-- 1129
      K V I+ ++ IR+A+EV+ Q + L +L + SFV+ +G +N
Sbjet: 497 KHVRIIRKLPIRDVAVREVLAQEL----DRPWKDLQVLRVAVMSFVRDFGPINHTTVS 552

Query: 1130 -----NRNLFQSDOKYSIASLEDESL--DPSGKSVITYKSLAFEKAL 1170
      N F D L+AS+ED L D+ I+T E+ +
50 Sbjet: 553 ITSDPESGETRESHRHNP/LQPPADFPQWLVASTEDYDLENATKAGFIPT-----ERV 607

Query: 1171 VPRPEKVKVYITALDAIWSLADGRQVDVAYNMSIYQVBSQMTLIEELQELIMPDPKYL 1230
      P V + +A DAL L + VD ++ + + + ++ HLG I DP
55 Sbjet: 608 SPFAPFV--ITSADALAVVNERGRVLDHIAELLHRDPD-DVVAELGSAITFRDP---- 660

```

-2128-

- Query: 1231 NGRMTYVSRCDFLSGDVVTKEVVDLFVKQDNQDFNWSHYAGLLEAIKPARITVADIDXR 1290
++ +LSG V KI+V + D ++ L ++P + +DI R
5
Sbjct: 661 -ADGSWQMDAYLSGSPVRDKIKVAAAAALDFV---YNNRVIALAGVQPDVLRPSDITAR 716
- Query: 1291 TGSRWIPLAVYKGFQAEITPAGKAYELSDQ-EVATVLEVSFIDGITVYQSKFATTYSNATD 1349
+G+ WIP A F +E M3 + E+A+ + G + A T TD
Sbjct: 717 LGAPWIPAADVVAFVKE-MMGTDRIHMHPELASWTVEARQLGYLA-----AGTSEWGTID 770
- Query: 1350 RSLGVPSASRYDSGRKIPENILNSNQPTITKQVVGDKKKNNVDVEKITVLRAKETHLQRL 1409
R ++ + LNS P I + +GD ++ V +V T K + +
Sbjct: 771 RR-----HAGELLSDALNSRVQIPDITRDGDSERRVLNVVITEAAKEKLHKIKDA 821
- Query: 1410 PQGFVAKYFEVQOMIEDTYNRLYNRTVSKSYDGSHLTIDGLAQNISLRPHOKNAIQRIVE 1469
PQ ++ P+ + YN +N + + G HL + G + L HQK I RI+
Sbjct: 822 PQRWINSDFDRTDRLARVYNDRFNNIAPRKFSGDHNLPGASGAFVLVGHQKRGITWIKIS 881
- Query: 1470 EGRALLNEHVSGSKTLMGLGAGFKLGLMVKHPLVVPVSSLTAFQGEIMKFFPTKKVY 1529
LAH VG+GKT+M + + LG++ K + VVP AQ +E + +PT ++
20
Sbjct: 882 SGSTYLARHVGAGKTMWMAASIMEQRFILIAKAMQVVGCHLAQAAREFLALNYPTARIL 941
- Query: 1530 VTTKKDFAKAKRKQFVSRITIGDYDAIVIGDSQFEKIPMSREKQVITYINDKLEQLREIKL 1589
V + +F+K KR +P+SR T +DAI+I S F I + + I-D+LE + L
Sbjct: 942 VADETFSKDKRRLSPRAATATDAIITHSAPRFIGVPAFESQMIHDELELTETLLL 1001
- Query: 1590 GSDSDYTV--KEAERSIKGLEHLEELQKLERDTFIEFNLGIDFLVDEAHFKNIRPI 1647
+ + V K ER +GL+ +LE L +D + +G+D + VDEA F + +
Sbjct: 1002 KVEDERDSRKRERLERKEGLQERLEAI-ST-RKDDLTIAEIGVDQIIVDEAQEFKLSFA 1060
- Query: 1648 TGLGNVAGITMTTSKKNVDMNMKVRQVQAEHGDNRNVFAITGVPSNISSELFMTMDYIQP 1707
T + + G+ S++ D+ +K R ++ + R +V A+GTP++N++ E+F++
Sbjct: 1061 TWSITLKGVDVWNGSQRANDLVKSRFLETINFGRALVLAAGTPIITNLGEMPSVQRVIMGH 1120
- Query: 1708 DVLERLYVNSDFSWGAFGNIEISMELAPTDGKYQPKRKKFVNLPELMRIYKSTADI- 1766
LE + + FD+W FG+ +EL P+G KY+P RF FVN+PEL+ ++ AD
Sbjct: 1121 AALEKREGLRHFDAWASTFGDTTTELELOPSG-KYKFSVRFAFVNVELIAMFERSFADVV 1179
- Query: 1767 ---QTSMDLDF-VPEAKIIAVESELTQAKQYKLEELVKRSDAIKSGS--VDPSSRDMKL 1820
+ + +P + + V S+ TQA K++ L+R AI+ P D +
40
Sbjct: 1180 MPADLETVKVPALSTGRROIVTSKPTQAFKHQMVLAEIRIKAEERERPPQGGDILLS 1239
- Query: 1821 ITGEARKLAIIMRLIDPTYSLSNDKILQVVDNVERIYRDGAGK-----AT 1867
+ + R AID+RL+D + K+ +V N RI++ AG
Sbjct: 1240 VITDGRHAAILDLRLVDADNDNEPDNKILNLVNSAFRIWATAGSVILRHDSKPFVEPGA 1299
- Query: 1868 QMIFSDIGTPK-SKEBGFVYNELKDLFVDRGPKREIAFVHDANTEKKNLSRKRVNS 1926
QMIFSD+GT K GF Y +D+ +G+P ELAF+ D E K L V +G
Sbjct: 1300 QMIFSDLTGISVEKTRGFSAYRWIRDELIRLGVPASEIAPMQDFKSGQRPLGQVRAG 1359
- Query: 1927 EVRILMASTREKGGTGLNVQSRMKAVHYLDVMPRPISDIVQRNRLIRQNNHQEVDIYHI 1986
VR L+ +S+E GTG+NVQ R+KA+H+LDVW PS I QR GR+RQNH H EVDI+ Y
Sbjct: 1360 RVRFLIGSSKTMGTGVNVQLRLKATHHLDVFWLFSQIBQRGRIVRQNHQEVDIYATA 1419
- Query: 1987 TKGSPNYLQATQENKLGKYYITQIMTSKDFVRSADIDE-QTMSADFALATGNPYLKK 2045
T+GS D +WQ E K ++I ++ +R EDI E Q + K++G+ L K
Sbjct: 1420 TEGSLDATWQNNKRRKAFIAAALSGDTSIRRLREDIGEQANQPMWMAIASGQRLMK 1479
- Query: 2046 MELENELTVLENQKRAFNRSKDEYRHTISYSEKHLPIMKRLSQQYKDIAQSLATKSQDF 2105
LE ++ LE + A + R + +E+ + +R+++ +DI + + T +DF
Sbjct: 1480 AGLSADIARLERLEAAHIDDQHAVRQLDAERDIEVSTRRIARIGQITRLVIVTTGDF 1539
- Query: 2106 VMRFNQDMQNMRAAGDYLRK-LITYNRSETKVEKVTASFRGFOLKM-TTRGASRPLEET 2163
M + + R EAG L K ++T + + +AS GP+L+ R + T
Sbjct: 1540 TMTVAGKDYSEKKEGRALMKRILTLVLSPEGEAVIASIGOFFLEYHQRYGKDYRYT 1599
- Query: 2164 ISLMIVGDNQYTVALLDK-SDWGTIQRISNAIDHIDDQKTELKVLKDKLRVAKVSV 2222
L G + Y + L + + G + R+ +A+D +E+ + + +D + L +
Sbjct: 1600 TMLKRTGAD-YELIFVIVTFLGAVSRILSHALODDFDKERKRYVRQGDARRRLASYSQSG 1658

-2129-

Query: 2223 DKVFPKEDYQLVKAKYDVLAPLVEKEASIEEIDALIA 2260
 + + + L EK ++ E++ ALA
 5 Sbjet: 1659 E-----GSEFAPAGELAEKKHRLAEVETALA 1684
 Identities = 99/271 (36%), Positives = 153/271 (55%), Gaps = 10/271 (3%)

Query: 607 RDKVETNIVAIRLVKNLSEVHRNASPSEQELIAKYVWG--LANEFD-----DYNPKF 659
 +D+ NI ATRL +E R A+ EGE L ++ G+G LAN F ++ +
 10 Sbjet: 80 KDRARUNIAAIRLAAEIRASKRPATRESQETLIRFTGFGASDLANGVFRPGELEPRKGN 139

Query: 660 SKERSELKSLVTDKEYSDMKQSLTAYYTDPSLIRQMWDKLERDGTGOKILDPEMGTGN 719
 + +L+ V + +Y+ + + A++T ++R +W L+R G+ G++L+P +GTG
 15 Sbjet: 140 DEIGSLELDVAGSTDYASLARCTQYAHFTPEFTVRAIWSGLQRLGMRGRVLEPGITGL 199

Query: 720 FFAAMPKHLREKSELYGVELPTITGAIKHLHFNHSHIEIKGFETVAFNDMSFDLVISNVE 779
 F A MP+ LR+ S + GVELD +T I + L P + I F SFDL I N P
 20 Sbjet: 200 FPALPMEALRDLSHTVGVELDFVTCIVRLQPRARILTGDFARTEL-PASFDLAINPP 258

Query: 780 FANIRIALNRYDRP--YMHDFVKKSLDLDHGGQVAIISSTGMDKRTENILQDIRET 837
 F++ + +R R +HDYFV +S+DL G A ++G+GMDK Q I T
 25 Sbjet: 259 FSDRTVREDRAYRLGLRLHDYFVARSIDLKFGAFAPVTSSTGMDKADSAARQHIAT 318

Query: 838 TEFLGGVRLPDGAFKAIAGTSVTTDMLEFFQK 868
 + + +RLP+ +F+A AGT V D+LFF+K
 30 Sbjet: 319 ADLIAAIRLPGSGFRADAGTVVVVDILFPRK 349

SEQ ID 5856 (GBS327N) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 148 (lane 8-10; MW 140kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 148 (lane 11-13; MW 115kDa) and in Figure 182 (lane 8; MW 115kDa).

Purified GBS327N-GST is shown in Figure 243, lane 5; Purified GBS327N-His is shown in Figure 235, lane 5.

GBS327C was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 148 (lane 14; MW 73kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1882

A DNA sequence (GBSx1990) was identified in *S.agalactiae* <SEQ ID 5857> which encodes the amino acid sequence <SEQ ID 5858>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3656(Affirmative) < succ>
 45 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2130-

Example 1883

A repeated DNA sequence (GBSx1991) was identified in *S.agalactiae* <SEQ ID 5859> which encodes the amino acid sequence <SEQ ID 5860>. This protein is predicted to be giant membrane protein. Analysis of this protein sequence reveals the following:

```

5   Possible site: 33
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10   bacterial cytoplasm --- Certainty=0.3698(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:AAG19662 GB:AE005054 calcium-binding protein homology; Cbp
    [Halobacterium sp. NRC-1]
    Identities = 22/43 (51%), Positives = 29/43 (67%), Gaps = 1/43 (2%)

    Query: 9 KSDQDGLTDAELAL-GTDFQSVDTYGDQADLELQSGHSP 50
    +D+D DGL+D E+ + GTDP DTGDD G EL+G P
20   Sbjct: 198 RDTDDGLSDGVEVRVAGTDPFTRDTGCGVDDAELRAGSLP 240

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1884

A DNA sequence (GBSx1992) was identified in *S.agalactiae* <SEQ ID 5861> which encodes the amino acid sequence <SEQ ID 5862>. Analysis of this protein sequence reveals the following:

```

    Possible site: 52
    >>> Seems to have no N-terminal signal sequence

30   INTEGRAL Likelihood = -2.39 Transmembrane 1609 -1625 (1609 -1625)
    INTEGRAL Likelihood = -1.81 Transmembrane 30 - 46 ( 29 - 46)

    ----- Final Results -----
35   bacterial membrane --- Certainty=0.1956(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

    LPXTQ motif 1600-1604

```

The protein has homology with the following sequences in the GENPEPT database.

```

    IGB:X57841 antigen I /II [Streptococcus sobrinus] (v...

40   >GP:CRAA0973 GB:X57841 antigen I /II [Streptococcus sobrinus]
    Identities = 419/1436 (29%), Positives = 608/1436 (42%), Gaps = 310/1436 (21%)

    Query: 23 KSKKYRTLCSVALGTWVAVVAMGGTVAHAEVTVSV-----DITYQRTS--NPAITLPEA 76
    K K RTL LST + A A G A A+E +T+ DT + TE NPAITLPE+
    Sbjct: 23 KVKSGRTLSGALLGTATLASGA--GQKALARETSTSTSTSGGUTAVVGTTGNPATLPPK 80

50   Query: 77 QPNP-----VSEQTESMASTGQNGRIAVTVPHDTVT-----QAVE 112
    Q NP V T + +8 VTV D + +
    Sbjct: 81 QNPSSQAETSQAQARQKTGMSVDVSTSELDRAAKSPQEAQVTVSQDATVNGTSPD 140

    Query: 113 EAKAEGSVTVEDSPMDLGNTRSAVET-----NQQIS-----K 144
    EA + +D + + A E NQ+I+ F
55   Sbjct: 141 EANKREPEITKDDYSQAADICKATEDYKASVAAMQASTDRINCELAAKKAQYEQDLAANK 200

```

-2131-

Query: 145 AD-----ADTKQVSTINRVTX----TYKADKATYESNKARISEQN 181
 5 A+ A QK + I + Y A K Y+ AR++ N
 Sbjct: 201 AEVSRLMRMKRPRIYEAKLAQNKDLALAIQANSDSQAYAAAKRAYKQWVRQJAN 260

Query: 182 KELSQAIEGANQTGKETNAWVDTKVNLDKARYADADVTVKRQ-----VVSNGNQTSLV 234
 +AVE A N + ++ ++ R A AD K +GN +
 Sbjct: 261 AAAKAYBERALAAANTAKNDQIKARIRAIQQRSAKADYRAKLQYBKDLAAQAGNAANE 320

Query: 235 DY---TNYGKAVETIQSTNEQAVADY---LTKKTKADDIVAKNQAQKNEA----- 280
 DY Y + + +Q+N N A Y K I A+N+AIQ+ +A
 Sbjct: 321 DYQAKKAYEQELARVQORANAAKQAYEQALANSAKNAQITAEAEAIQQAQAKADYEA 380

Query: 281 -----GLANAKADNEAIERRNQAGQAANDAE--RAGQAANDQANQEQKQVLSDRAA 330
 LA A++ N A E Q AA + E +A AA QA ++ Q + + A
 Sbjct: 381 KLAQYQKDLAAAGQNAANEADYQEKLAAYEKELARVQORANAAKQAYEQVQANAKNA 440

Query: 331 EIEAITKRNKEKRAARKNEAIDAYITKEMERQORDLAEIS----- 372
 EI + +E+ A A + E + +E+ +Y++DLAE
 Sbjct: 441 EITENRAIRKERNAKFTDYELKLSKYQELAQYKDKLAEYPAKLQAYQDEQAIIKALA 500

Query: 373 ----KGEETIASELAQALNLNNGEPQAGHAIIRN----- 404
 K E+G +SE AQ+L + + EP AQ +T
 Sbjct: 501 ELEKIHKNEDGNLSEPGAQSL-VYDLEPNAQVALVTDGKLLKASALDEAFSHDEKYNNGH 559

Query: 405 --PDQI-----ISTGDALLGGYSRIIDSTGF-----FVYDMFKITGELS 441
 PD + +++ L G + D G+ F + K G++
 Sbjct: 560 LQPNLNVTYLBQADDVASSVELEFGNG---DKAGWTTVTSNGAEVKFASVLILKRGQNT 616

Query: 442 FNYQNQLHARPDGKKISRVTYDITNLVSPAG-----TNAVKLVPNDPTGFIAYRNDGN 496
 Y NL+++ +GKKIS+V Y T V P T V L + DPT G A G
 Sbjct: 617 ATYTNLKQISYNGKIKISKVYKCT--VDPDSKFNPTGNVNLGIFTDPTLGVFASAYTQQ 674

Query: 497 GDWRID---KMEPRVAKYYLEDGSGVTFPSKEKPGVTHSSHLNHDIGLEYVKDSSGKVF 553
 + T K EF +Y EDG+ + + +SLN +E KD SG VF
 Sbjct: 675 NEKUTSIFIKNEF---TFYDEGNDPFDN---ALLSVASLANREHNSIEMKADYSGTVF 727

Query: 554 PINGSTVQVFN-----EGLARSLGSRASDNLNLPSEWDTTSSRYAYKQAV 599
 I+GS++ N EG + RAS+ WD+ + ++ QA
 Sbjct: 728 KISGSSIGEGKNMIYATDTLNFKKREGGSLHMTYTRASEPG--SGWISADAPNSWYGAGA 785

Query: 600 STVTSQNTY-----TVTFGGQDMPQNVGL-----SYWFALN----- 630
 ++ N Y T +MPQ G + W++LN
 Sbjct: 786 VRMSGPNMYITLGATSATNVLSLAEMPQVPGKNDNTAGKKPNINWYSNGKIRAVNVPKVK 845

Query: 631 --TLFVARTVTPSPKHVIVEL-----EPIPEPITVTPDIYTPKFTPEKPVITF----- 679
 P P P V EL EP EP TP P PEKPV T
 Sbjct: 846 EKPTPFVEPTKPEPTYEVEKELVDLPVBPKYEP-EPTPEKNPDQSIPKEPVBPTPYEVE 904

Query: 680 ----PKPLDEVQPSLTLTKVT-----LPVKPIPEKLPTP-----QVPTV 716
 P P++ + T + T PV+P + LPTP VPTV
 Sbjct: 905 KSLPAPVPEPSYKRPPTPPQSTPDQKEPEKPEVPSYQSLPTPFVEPVYETVPGVPSVPTV 964

Query: 717 HYHAYRLTTSHIMKEVVNSQANLHEKTVAKDSTVIYELTVDALSPNRAQTSLIFEDY 776
 YH Y+L + KB+ N D ++ + VAK STV + L L R +TTS + D
 Sbjct: 965 RYHYTKLAVQGVTKIKNKQDDLDIDKTLVAKQSTVKFQLKTADLPGRPETTSFVNDP 1024

Query: 777 LPAGYLEFKETIQKENGNYVLSFDEKTNFVTLAKENLQBVNKDLQVQVLTAPKLYGS 836
 LP+GY + E T+ + + S+D + VT TA L +N+DLT+ P + G
 Sbjct: 1025 LPSGYQLNLKATKVASPGFEASYIAMHTVTFATAETLAALNQLTKAVATYTPVVGQ 1084

Query: 837 VQNDGATYSNSYKLLKNGKTTNAYTVTSNVVIVRTPG-----DGETTLLTPDKKNENAD 891
 V NNGATY+N++ L++N +AY + SN+V V TKG D + ITP K N++
 Sbjct: 1085 VLNDGATYTNFNTLMVN---DAYGIKSNIVRVTTGKPKNDPDPNSNYITPHKVIKHEH 1140

Query: 892 GVLINDTVVALSTINHYKLIMDLDDQYKEDRSAKETIARGFFVDDYPREVLVDVNGTAI 951
 GV+I+ V GTN++ LTMDDQYKGD+SAKE I +GFF+VDYPER LD+ + +
 Sbjct: 1141 GVVLDGKSVLASTINHYKLIMDLDDQYKGDKSAKEIKGFFVDDYPREALDLRLTLIKL 1200

-2132-

5	Query: 952	TTLDGQKVSIGITVKNYASLNKAPKUIQDKLARAKITPTGAPQVFMFDNQAFYDQVQVG 1011
	Sbjct: 1201	TDANGKAVTGVSVADYASLEAAPAAVQDMLKANIIPKGAQVFTADDPQAFYDAYVVG 1260
10	Query: 1012	TSLALLTMTVNDKSLYQGTETYNKAYQVDFNGNYETKEVTINTLVSPEPKKQ-NNLNDKV 1070
	Sbjct: 1261	TDLTIVTMTVKAEMGKTGGSYENRAYQIDFGNGYESNLVNNVVKINPEKDVITLNDIFA 1320
15	Query: 1071	D---INGKPMVGTQNHYYLTDLDQYRGIKADNSQIAQGFYVDVYPE-----EALLPD 1122
	Sbjct: 1321	DSTNVDSQTTALNQVFNRYLIGGI-----IPADHARELFYYSFSDDYDQGTGXQYQVKA 1375
20	Query: 1123	EAAIQFVTSDGKIV-SGITVKSY--SQLEAPKTLQAASFQKIQPKGAQVFMPE 1175
	Sbjct: 1376	FAKVDLTLDGTLIKAGTDLTSYTEAQVDEANGQIVVTFKEDFLRSVSVDSAPQAE 1431
Identities = 209/442 (47%), Positives = 280/442 (63%), Gaps = 27/442 (6%)		
25	Query: 1198	TVLETMNSGKSY-ENVAYQVDFGQAYETINTVINFVPK-----VTPHKSNTNQ 1244
	Sbjct: 1080	TVVGQVLNDGATYTNNFTLNVDAYGKISNIRVITPGKENDPNDNSNYITPHKVNKIE 1139
30	Query: 1245	EGISIDGKTVLEPNTVNYKIVLDYSQYKDMVVTDVLAKGFYVVDVYPEALTLNPDGQI 1304
	Sbjct: 1140	NGVVDGKSVLAGTINYELTMDLQYKSGKSAKEIQKGFYVDVYPEALDLRLTOLIK 1199
35	Query: 1305	VLDKDNKRVSGISVSTYASLSEAPKVVQDAMAKRQFTPKGAIQVLSDDPKFYDYTVKT 1364
	Sbjct: 1200	LTDANGKAVTGVSVADYASLEAAPAAVQDMLKANIIPKGAQVFTADDPQAFYDAYVVT 1259
40	Query: 1365	GQTLVTVLPMTVKNELTKTGGQENTAYQIDFGLAYVTEVNNVVKLPQKQVVDLDSH 1424
	Sbjct: 1260	GTLDTIVTMTVKAEMGKTGGSYENRAYQIDFGNGYESNLVNNVVKINPEKDVITLND 1319
45	Query: 1425	KDA-SLDGKEVALHQTFNRYLWGMIPSNRAITDLFEYGFEDNYDEKHDEYNGVYSYLT 1483
	Sbjct: 1320	ADSTNVDSQTTALNQVFNRYLIGGIIPADHARELFYYSFSDDYDQGTGXQYQVKA 1379
50	Query: 1484	DVILKDGKSVLKEGTEVTKYTLQVVDTEKGLVLSIFPKSFLETVSDSAPQADVYLQMKRI 1543
	Sbjct: 1380	DLTLKDGTLIKAGTDLTSYTEAQVDEANGQIVVTFKEDFLRSVSVDSAPQAFVYLQMKRI 1439
55	Query: 1544	AAGQVENTYLRHTVNGYVSSNTVTVHTPQPEEPSNQ-----TPQPPIETIEPPV 1595
	Sbjct: 1440	AVGTANTYVNTVNGITYSSNTVITSTPEKQSPVDPKTTITVFPQPRGKAYQAPPA 1499
60	Query: 1596	PASILENTGQES----LLGLI 1613
	Sbjct: 1500	GAQ-LPATGDSNAYLPLGLIV 1520
Identities = 100/210 (47%), Positives = 137/210 (64%), Gaps = 4/210 (1%)		
65	Query: 1060	PKKQNLNKDKVDINKPMVGTQNHYYLTDLDQYRGIKADNSQIAQGFYVDVYPEAL 1119
	Sbjct: 1132	PHKVNKENGVVDGKSVLAGTINYELTMDLQYKSGKSAKEIQKGFYVDVYPEAL 1191
70	Query: 1120	LPDEAAIQFVTSDGKTVSGITVKSYSQLRAPKYLQAASFQKIQPKGAQVFMPEDPQA 1179
	Sbjct: 1192	DLKTLIKLTDANGKAVTGVSVADYASLEAAPAAVQDMLKANIIPKGAQVFTADDPQA 1251
75	Query: 1180	FFESYVTKGENITVITPMTVLETMNSGKSYENVAYQVDFGQAYETINTVINFVPKVTPHK 1239
	Sbjct: 1252	FYDAYVVTGDLTIVTPTMTVKAEMGKTGGSYENRAYQIDFGNGYESNLVNNVVKINPEK 1311
80	Query: 1240	SNT-----NOBGISIDGKTVLEPNTVNYKIV 1265
	Sbjct: 1312	DVTLTMDPADSTNVDSQTTALNQVFNRYL 1341

There is also homology to SEQ ID 598.

-2133-

SEQ ID 5862 (GBS76) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 2; MW 17.4kDa). The GBS76-His fusion product was purified (Figure 196, lane 8) and used to immunise mice. The resulting antiserum was used for FACS (Figure 294), which confirmed that the protein is immunoaccessible on GBS bacteria.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1885

- A DNA sequence (GBSx1993) was identified in *S. agalactiae* <SEQ ID 5863> which encodes the amino acid sequence <SEQ ID 5864>. This protein is predicted to be abortive infection bacteriophage resistance protein (abiEi). Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
15 bacterial cytoplasm --- Certainty=0.2765(Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

- A related GBS nucleic acid sequence <SEQ ID 9931> which encodes amino acid sequence <SEQ ID 9932> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

- ```
>GP:AAB52382 GB:U36837 AbiEi [Lactococcus lactis]
Identities = 51/206 (24%), Positives = 90/206 (42%), Gaps = 23/206 (11%)

25 Query: 17 KNNIGIVNCKKALGIFTYILTRLEKEGIIIFRVEKGIFLTQNGDYDEYFFQYRFFKAIIF 76
 K G + K + GI YL + + + V+KG+++ + D + FQ ++ KA+
 Sbjct: 76 KYKGNIIIRKIVRDEGISDYILRKFLKYNLTAEVDKGVYIFPHKKKDSLPFQKQYSGAVI 135

30 Query: 77 SYISALYLCQFTDEIRQYFDVTVPRGYF-----NTPPANLNI 114
 S+ ++LHLQ D IRQ ++VP Y H N+ I
 Sbjct: 136 SHETSILYLCQVDYIPIQKIQSVPEKYNISRIQEPHENLTSYNYVDINSINIMDGNIPT 195

Query: 115 HFV-SKEYSELGNTTVPTFGANNVRYVDPERIICDPVHIREKIDSELVKTLSQSYQNYFK 173
 + V +K S + TV + +G +RV R I D + K + E+ + + Y
35 Sbjct: 196 MLVRNKSISPTQIETVNSFLGLPLRVTSIARSIVDLKPSHKAEEVKEQAIIKYLERFF 255

Query: 174 KNLAKLYEYATQNTLEKVKQTLLEVL 199
 N+ +L A N L+++ L +L
40 Sbjct: 256 DNIVRLKRIAKTQVVLKLEYLYLIL 281
```

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1886

- 45 A DNA sequence (GBSx1994) was identified in *S. agalactiae* <SEQ ID 5865> which encodes the amino acid sequence <SEQ ID 5866>. This protein is predicted to be abortive infection bacteriophage resistance protein (abiEii). Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence
50 INTEGRAL Likelihood = -1.12 Transmembrane 260 - 276 (259 - 277)
```

-2134-

----- Final Results -----

bacterial membrane --- Certainty=0.1447 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAH52383 GB:U36837 AbiRii [Lactococcus lactis]

Identities = 76/276 (27%), Positives = 135/276 (48%), Gaps = 19/276 (6%)

Query: 14 SKNGLTFNSVMTYTFLEVLKLLSQSSSYNHYIFKGGFLNINVGVSRSSTDVIDFLPH 73  
 ++N + + Y E L +LS S Y ++ KGGFL+ + R+T D+D  
 Sbjct: 12 TRNDDIGIENNRIRIYATERFLRLSASQYKEKFLKGGFLIGVTYNSQRTTKDLITALI 71

Query: 74 QITLSESTVKQQLKEIL-ADSESGISFVIQSITTIKESDYGGYRATISOQLE--NIKQV 130  
 +++++ + EI D E+ + F ++ +T+ ++ Y GYRA + N +  
 Sbjct: 72 DFKSDAQSIERVITEICNIDLEDQVLFLKELTSSQDMRLIYPGYRAKLKNNPFDGNTRID 131

Query: 131 IHLDIATGDEVTPQIPITYDYKAIFDE-----DNFPIIAYITKITLAELQITYSRNFNS 185  
 LDI GD +TP+ IF+E ++AY ETI AEKL+TI +R +N+  
 Sbjct: 132 FDLIDGVDGRTPEAKKIKIPLIFNEVKGVEKQIEVLAYPKETIQAELKLETILTRGVNT 191

Query: 186 RSKDFPDVYIL--SKLKKKIDFNQNLKNCQRTFSYRE-TELDFEKIE-----LLERFK 237  
 R KD+YD ++L + I F A +T+ +R T+ E++ E LE +  
 Sbjct: 192 RMCQYDFELLITDQENSNISISFY---AFKNWFEFRNPTQPIDELFEOWLFILDEILE 248

Query: 238 SDPTNQQWQWNYKKYSYTKGISLANVLDEMISLIT 273

S + + W NY K +Y K +++ ++ E+ ++

Sbjct: 249 SKLEKQWPNYKIDRNYAKHLNDDIISRIKEFVS 284

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1887

A DNA sequence (GBSx1995) was identified in *S.agalactiae* <SEQ ID 5867> which encodes the amino acid sequence <SEQ ID 5868>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1137 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1888

A DNA sequence (GBSx1996) was identified in *S.agalactiae* <SEQ ID 5869> which encodes the amino acid sequence <SEQ ID 5870>. Analysis of this protein sequence reveals the following:

Possible site: 44

>>> Seems to have no N-terminal signal sequence



-2135-

## ----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2782 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1889**

A DNA sequence (GBSx1997) was identified in *S.agalactiae* <SEQ ID 5871> which encodes the amino acid sequence <SEQ ID 5872>. Analysis of this protein sequence reveals the following:

15 Possible site: 21  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -10.14 Transmembrane 310 - 326 ( 301 - 334)  
 20 ----- Final Results -----  
 bacterial membrane --- Certainty=0.5055 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:AA038044 GB:AF295925 Orf28 [Streptococcus pneumoniae]  
 Identities = 272/344 (79%), Positives = 307/344 (89%)  
 Query: 568 VYVNPAPYFPKVIQVQTITLPTIQPGGDEPERAKAIYDLYKSGKATQAIILGNWSV 627  
 +YVNP FYTPKVIO+QTTILP IQPGGDEPERAK IY++LKS+GA+ QAIAAILGNWSV  
 Sbjct: 1 MYVNPQFYFPKVIQVQTITLPTIQPGGDEPERAKHIEFLKSGQAIAAILGNWSV 60  
 30 Query: 628 ESSINPKRABGDVLPSPVATDSWDEGWLTLNPTITNCRVFNILKRLGLGQNTDTA 687  
 ESSINPKRABGDVLPSPV MDE WL + GP IY+G YFNIL RGLGLGQNTDTA  
 Sbjct: 61 ESSINPKRABGDVLPSPVFPVPPWDESWLAIQGPATYSGATFNILHRLGLGQNTDTA 120  
 35 Query: 688 DGSRRHTLLLEYAKGHQKQVYDLGLQDLPMLHGDSPYYTNLKDFFPNKSGSPASIAQLFL 747  
 DGS RHT LL YA+ ++KMYDL LQDLPML+GDSPPY +WLKDPFN+GS A+LAQLFL  
 Sbjct: 121 DGSRRHTALLHARTQNKMYDILLQDLPMLHGDSPYYQSKLDPFNKSGSPASIAQLFL 180  
 40 Query: 748 TYMEGNSGDKLLERQTRASEWYQIEKGPSQPMGGTAQSDPKALEAVREDLFENSIPGG 807  
 TYMEGNSGDKLLERQTR+EWYQIEKGPSQ NOG A+SDP++LE VR DL+++S+PGGG  
 Sbjct: 181 TYMEGNSGDKLLERQTRATEWYQIEKGPSQTMGGQAQSDPQSLGVRGDLYDSVPGGG 240  
 45 Query: 808 DGMVAYGQCTGWVAARNQLGLKLGKNGRGIPIIITWNGQDWVTAASLGGEITGSP 867  
 DGM VAYGQCTGWVAAR+NLGLKLGKNGRGI II+TWNGQDWV T++SLGGEITG+P  
 Sbjct: 241 DGMVAYGQCTGWVAARNQLGLKLGKNGRGIPIIITWNGQDWVTAASLGGEITGSP 300  
 Query: 868 QEGAILSPAGGGRGIPTEYGHVAPVKEVYDPGGSFLSETNYGN 911  
 + GAI+SF GG HGTP YGHVAPVKEVY DGSFL+SETNY GN  
 Sbjct: 301 RAGAIIVSVQGTGTGPASYGHVAPVKEVYDUGSFLVSETNYGN 344  
 50

SEQ ID 5872 (GBS74d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 121 (lane 3 & 4; MW 95.5kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 121 (lane 5-7; MW 70.5kDa) and in Figure 179 (lane 9; MW 70.5kDa).

55 GBS74d-His was purified as shown in Figure 233, lane 7-8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1890

A DNA sequence (GBSx1998) was identified in *S. agalactiae* <SEQ ID 5873> which encodes the amino acid sequence <SEQ ID 5874>. This protein is predicted to be TrsE-like protein. Analysis of this protein sequence reveals the following:

Possible site: 55  
>>> Seems to have no N-terminal signal sequence

```

10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5526(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

15 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG38042 GB:AF295925 Orf26 [Streptococcus pneumoniae]  
Identities = 618/782 (79%), Positives = 712/782 (91%), Gaps = 1/782 (0%)

Query: 1 MKKLKISMKSK-TSSNDKKQKTKTQKEISPSTVNTLAYQGLFONGLMQVSPSYFSQT<sup>YL</sup> 59  
MK+ +++K + TS+ +KK++ K +K+E+ PST NTL+YQ L+QNGLMQV YFSQ+YL  
Shit.: 2 MDEKENTLKKGGCTTTWKPKRETKPKREU.DGTAHT.SVALYONGMLMOUKEDYRSGEV. 62

Query: 60 LGDVNYQTIVGLDDKGAIVKYSIDLINSLDDKTNFQLTIFNQKNLEKFRKSIILYPLQEDG 119  
 LGDVNYQTIVGL+DKGAI+EKYSDLI SLDD+TNFQLTIFN+++NLEKFR S+LY +EDG  
 25 Spict: 63 LGDVNYQTIVGLEDKGAIIEKYSDLIKSLDDDTNFQLTIFNKRNLNLEKFRHSVLVEEKEDG 122

Query: 120 FDTYRDELNRNM+DANLEAGENNFSAVKFLSFGKSDQTPKLAFRSLSQIGEYFKSGFSEID 179  
+D+YR ELNRNM+ NL++GENNFSAVK +SPG+ D PK A+RSLSQIGEYFKSGFSEID  
Sbjct: 123 YDSYRKELNRNMNQNLDSGENNFSAVKLISFGKSDNPKQAYRSLSQIGEYFKSGFSEID 182

Query: 180 VSLGLLGGEEVRNVLADMLRGENHLPFSYKDL/TLGSGSTKHFIAPTYLSFKHKNHIELD 239  
L GEERVN+LADMLRGE+HLPFSY+DLT SQQ+T+HFIAP L FK+KN++++D  
Sbjct: 183 ARFESLAGEEVRNLLADMLRGEHLPFSYRDL/TRSGQTTRHFIAPNLLDFKNKNYLQIND 242

Query: 240 RLLQIVYVRDYGMELGDKFIRDLMQSDLEVMISLHAGSSTKSETWTKLRTKTKTLMESQKI 299  
RLLQIVYVRDYGMELGDFIRDLMQ DLE+++SLHA+ STKS+ M KLRTKTKTLMESQKI  
Sbjct: 243 RLLQIVYVRDYGMELGDFIRDLMQGDLRLIVSLHAQSSTKSDAMKKLRTKTKTLMESQKI 302

40 Query: 300 GEQQKARTGIYLEKVGHVLENNIDEAEALLKTMTETGDKLFDVTVFLIGVFGQDEELKQ 359  
GEQQK+ARTGIYLEKVGHVLE+NIDEAE LL+TMT+TGDKLF TVFLIGV E++LKQ  
Sbjct: 303 GEQQKLRATGIYLEKVGHVLENNIDEAEELLKTMTETGDKLFDVTVFLIGVFGQDEELKQ 362

45      +LD ++QVAGSND++ID L YMQRAAFNSLLPFG ++LEQVSRSLTNSIAVNPWTSVD  
 Sbjct: 363 ALDTVQVAGSNDLMIDKLPYMQRAAFNSLLPFGCDFLEQVSRSLTNSIAVNPWTSVD 422

50  
+ D+ GK+YGINQISSNII+IDR LNTPSGLILCTSGAGKGKMATKHEII+TK+KE+ +  
Sbjct: 423 LQDRSGKYQGINQISSNII+IDRSLNTPSGLILCTSGAGKGKMATKHEIITTKIKESGEN 482

Sbjct: 483 TEIIIVDPEAEYSVIGRTFGGEMIDIAPDSEYTLNVLDLSEENMDEDPVKVKSEFLLSFI 542

Sbjct: 543 GKLLDRKMDGREKSIIDRVTRLTYQSFKEPSLEEWVFLSQQPEEEAQNLAIDMELYVEG 602

60 SLDIFSH+TNI+T S+FLIYNVKKLGDELKQIALMV+FDQIWNRVV+NQKLGKKTWIIYFD  
Sbjct: 603 SLDIFSHKTIQTGSNFLIYNVKKLGDELKQIALMVVFDQIWNRVVRNQKLGKKTWIIYFD 662



[illegible]

The protein has homology with the following sequences in the databases:

Score = 427 bits (1086), Expect = e-118  
Identities = 257/785 (32%), Positives = 431/785 (54%), Gaps = 28/785 (3%)

Query: 9 DKQKTKTKQKRIK-----PSTVN-TLAYGCLPONGLMQVSSPSYSQTLVGLDW 56  
+K +T +Q +RI P T +Y + +G +VSP FS+ D+  
Sbjct: 11 EKKTKRACIKEDIVAKYKGGEDGTAQSPYVYVGVGKGVKSPKSCIEFAD 70

Query: 57 NYQTGLDDKAIGVEKYSDLINSLDKTNFQITIPNKNVGVGCKFRKSLIPLQLQEDGPDY 116

-2139-

+YQ D + AI KK DL N + D + Q + N + KV + ++ KS Q D F D  
 Sbjct: 71 SYQLAQPDQTATPEKLCDLNYYVDASIHQSFPLNRKVDVPQYAKSFELAPQDDDDI 130

5 Query: 117 RDELNMMMDANLEAGENNPNFAVKFLSPGKSDQTPKLAFLSLQIGEYFKSGFSDVSLG 176  
 R E ++ L G N K + L + F ++ K A L + I G F + +  
 Sbjct: 131 RAETTGIIQKQLANGNMGMVKTYLFTTTRAESVKAARARLKRIGFDLLGLYFKSGMAVAH 190

10 Query: 177 LLGSEBRVNLADMLGRNHL - PFSYKDLTSLGQSTKHFIAPTYLSFKHKNHRLDRLL 235  
 ++ G ER + N + L + + + F + K L SG STK FIAP + L F + + +  
 Sbjct: 191 VMDGWERLNLHGVIYHPDGEIFNFDWKLAPSGLSSTKDFIAPSSLCFNGAKTFMGSGKYG 250

15 Query: 236 QIVYVRDYGMELGDKPIRDLMSQSDLEVMISLHAKSGTSKSTMTKLRTKTKLMSQKIGBQ 295  
 + + + + EL D + D + + + V + + + LH + + + + ++ K T + + + KI BQ  
 Sbjct: 251 AVSFIQILSPSLSDMLADFLNTESGVLYNLHVQAIEQTKRAIKTIKRAITLDDAMKIASQ 310

20 Query: 296 QRMARTGIYLEKVGHEVLNNIDBARALLQMTYTKDGLFDTVFLIGVLADTEBQLKQSLD 355  
 + K R + G ++ + L ++ A + LL + ++ LF FL + + ADT + + L +  
 Sbjct: 311 KAVRSGYDMDLPSDIATYGEDAKKLLTKLQTRNRLFLQLTFLVLNADTKQKLNNDVF 370

25 Query: 356 IINQVAGSNMIDNLTVMQEAAPNSLLPFGKMYLEGVSRLLTSNIAPNAPWTSVDIHD 415  
 VA ++ + L Y QE S L F G N ++ + RSL TS + + AV P + + +  
 Sbjct: 371 QAAGVAQKGNCFPLVRLDYQQBQLGLASSLPLGVNQIK - IQRSLTSSVAVFVFVFTQELFQ 429

30 Query: 416 KGGK - FYGINQISSNIISDRGLNTPSGLLIGTSAGRGMAKTHEIISTLKLEADSDTE 474  
 G + YGIN S N + I + DR + P + L LFT G + GK M + K EI + S L D +  
 Sbjct: 430 GGAAMYTGINAKSNMIMDRKQARCFWALKLTGPSGKSMCKSEIVSVFTTDPD - - D 486

35 Query: 475 IIVDPENEYSIIQAGGESIDIAPDSTTFMLNLELS - DENMDEDPFKVKESEPLLSWIG 533  
 I I DFE EY + + G + I + + P S F + N L + + + + D + + P + + KS + F + LS +  
 Sbjct: 487 IFISDPEASYPLVKRLHSGVIRLSPTSKDFVNPLDINLNYSEDDNPLAKLSDFVLSPCE 546

40 Query: 534 KLLDRK - - MDGREKSIDRVTRLTYKHF - - - - - DTPSLVEWVFLSQQPSQEQANDAL 584  
 ++ K ++ EK + + IDR R + Y + + + P L + L Q BA + A  
 Sbjct: 547 LVMGKKGLEATKNTVIDRAVRVITYRPLADPFPBNPILSLDLKHALQCHVFEADRVAQ 606

45 Query: 585 DMELYVSGSLDIFSHRMTIKTDSHFLYINVKLODELHQIALAVIPDQIVSRVVMQKLG 644  
 ++LYV GSL + + F + HRTM + + + + K + LQ + LK + + + + + DQIV RV N + G  
 Sbjct: 607 ALDLYVSGSLAVFNHRTMVDIGNRLVSPDIKELGKQLKGLMLIVQDQIWGRVTANRSQ 666

50 Query: 645 KKTWYFDMLQLLLDLKVASDPFFKLHNRVRVKQAIPGTQVETVETLLDANGRIANS 704  
 K TW + DE LLL ++ + + + + W R RK + G IPTO TONY LL I + NS  
 Sbjct: 667 KATWYFADEPHILLKEEQTAAYSAEIWKPRUOGGIPTGATONVKLLSSPEIENTILNS 726

55 Query: 705 EPNKLLKQAKSDREELAVMLGLSKLEKYLVDPEKAGLIGKAGSTVVPFINKIPQHTGLF 764  
 + F + LQ QA DR + L L LS E + KY + N E G GL + + V + FP N IP + T + L +  
 Sbjct: 727 DFTILNQASGRKILAERLNLSTEQXYIDNSEPGEGLIFENVVLPFPNIPHTQLY 786

Query: 765 DIMST 769  
 IM + T  
 Sbjct: 787 KIMTT 791

SEQ ID 8926 (GBS75) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 11; MW 89.8kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 20 (lane 6; MW 114.7kDa).

55 GBS75-GST was purified as shown in Figure 197, lane 8.

GBS329 was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 8; MW 89kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 174 (lane 2; MW 114kDa).

GBS329-GST was purified as shown in Figure 220, lanes 9 & 10.

-2140-

**Example 1891**

A DNA sequence (GBSx1999) was identified in *S.agalactiae* <SEQ ID 5875> which encodes the amino acid sequence <SEQ ID 5876>. Analysis of this protein sequence reveals the following:

```

Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2442 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1892**

A DNA sequence (GBSx2000) was identified in *S.agalactiae* <SEQ ID 5877> which encodes the amino acid sequence <SEQ ID 5878>. This protein is predicted to be DNA-directed RNA polymerase ii largest subunit. Analysis of this protein sequence reveals the following:

```

Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4393 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1893**

A DNA sequence (GBSx2001) was identified in *S.agalactiae* <SEQ ID 5879> which encodes the amino acid sequence <SEQ ID 5880>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -9.92 Transmembrane 256 - 272 (250 - 277)
INTEGRAL Likelihood = -8.28 Transmembrane 216 - 232 (213 - 244)
INTEGRAL Likelihood = -8.12 Transmembrane 151 - 167 (148 - 191)
INTEGRAL Likelihood = -7.27 Transmembrane 57 - 73 (54 - 80)
INTEGRAL Likelihood = -6.74 Transmembrane 93 - 109 (88 - 111)
INTEGRAL Likelihood = -3.50 Transmembrane 172 - 188 (168 - 191)
INTEGRAL Likelihood = -2.76 Transmembrane 113 - 129 (110 - 130)

----- Final Results -----
bacterial membrane --- Certainty=0.4970 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

-2141-

>GP:AG38039 GB:AF295925 Orf23 [Streptococcus pneumoniae]  
Identities = 71/86 (82%), Positives = 83/86 (95%)

5 Query: 37 VESLADENITVMSYMTATYKIMQLGVAILAVVLLEFSKMAKKIANSQGAMTFEALAP 96  
+KSL+ +NPTVW+YM++ITK +MQPLOWAIL+VVL+LEFSKMAKKIANSQGAMTFEALAP  
Sbjct: 1 NKSLSYNTVMTWSSITKSMQPLGVAILSVLLEFSKMAKKIANSQGAMTFEALAP 60

10 Query: 97 MIVSYIMVAVVITMTTVIVEAIIAIA 122  
M++SYIMVAVVITMTTVIVEAII IA  
Sbjct: 61 MLSYIMVAVVITMTTVIVEAIIAIA 86

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### 15 Example 1894

A DNA sequence (GBSx2002) was identified in *S.agalactiae* <SEQ ID 5881> which encodes the amino acid sequence <SEQ ID 5882>. Analysis of this protein sequence reveals the following:

Possible site: 25  
>>> Seems to have no N-terminal signal sequence  
20 INTEGRAL Likelihood = -7.54 Transmembrane 32 - 48 ( 25 - 52)  
INTEGRAL Likelihood = -4.09 Transmembrane 63 - 79 ( 62 - 80)

----- Final Results -----  
25 bacterial membrane --- Certainty=0.4015 (Affirmative) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9933> which encodes amino acid sequence <SEQ ID 9934> was also identified. A related GBS nucleic acid sequence <SEQ ID 10777> which encodes amino acid  
30 sequence <SEQ ID 10778> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### 35 Example 1895

A DNA sequence (GBSx2003) was identified in *S.agalactiae* <SEQ ID 5883> which encodes the amino acid sequence <SEQ ID 5884>. This protein is predicted to be TrsK-like protein (traK). Analysis of this protein sequence reveals the following:

Possible site: 34  
40 >>> Seems to have a cleavable N-term signal seq.  
INTEGRAL Likelihood = -7.38 Transmembrane 66 - 82 ( 62 - 85)

----- Final Results -----  
45 bacterial membrane --- Certainty=0.3951 (Affirmative) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AG38037 GB:AF295925 Orf21 [Streptococcus pneumoniae]  
50 Identities = 343/457 (75%), Positives = 385/457 (84%), Gaps = 24/457 (5%)

-2142-

Query: 142 LIVIGSGGAGKITPRFVKFNLIQNCSENI VDPKDLARKTGKLFLENGYQVVKVLDLVNMT 201  
+ VIGGSGG+GKITPRFVKFNLIQ+N SNIVDPKDLARKTGKLFLE+GYQVKVLDLVNM  
Sbjct: 1 MAVIGGSGGSGKITPRFVKFNLIQNCSENI VDPKDLARKTGKLFLEHGYQVVKVLDLVNMK 60

Query: 202 NSDGFNPFRYIETENDLNRMLVYFNNTKGGSGRSDPFNDEASMTLVRAIASYLVDFPNP 261  
NSDGFNPFRY+ETENDLNRML VYFNNTKG+GSRSDPFNDRASMTLVRA+ASYLVDFPNP  
Sbjct: 61 NSDGFNPFRYIETENDLNRMLAVYFNNTKGGSGRSDPFNDEASMTLVRAIASYLVDFPNP 120

Query: 262 PGS-----SKQEQEARKKRGKYPAPSEIGKLIKLSKGNQDKS 300  
P + K+E E R+KGR F E + + + KS  
Sbjct: 121 PKTREQLMEESRLQKEQYQNLKQKQKVERRRKRSKPCESQNSINTYFVKTR-KS 179

Query: 301 ILEVLPHDYAKKYGHENFTMRNWADPONYKOKILDGVIATVTAKFALFNISQVILDTORD 360  
+LE+LPE+YAKKYG ENFTMRWADPONYKOKILDGVIATVTAKFALFNISQV+DLT+RD  
Sbjct: 180 VLEILFENYAKKYGHENFTMRNWADPONYKOKILDGVIATVTAKFALFNISQVMDLTOKRD 239

Query: 361 TMDLKTWOTQKTMVYLVIPENDITFRFLSAL-FFSTVFTLIRQADV-DPKQQLPIHVS 418  
T+D+KTWG+K+MNVYLVIPEND+TFRFLSAL FF+ FT + + + \*LP+HVR  
Sbjct: 240 TMDLKTWGQKESMNVYLVIPENDITFRFLSALLFTFNPYQTPNKTQSILMLRVLPLHVRV 299

Query: 419 YLDEFANVGEIPDFAEQTSIVSRNMSLVPIQLQIAQLQGLYKEKAMKTLGNCDLSLY 478  
YLDEFAN+GEIPDFAEQTSIVSRNMSLVPIQLQIAQLQGLYKEKAMKTLGNCDLS+Y  
Sbjct: 300 YLDEFANVGEIPDFAEQTSIVSRNMSLVPIQLQIAQLQGLYKEKAMKTLGNCDLSLY 359

Query: 479 LGGNDETFKFGSGLLQKQTVDSVTSRSFGQTGGSSSTSHQKIARDLMTADEVQNMGRDE 538  
LGGNDE+TFKFGSGLLQKQTV+DSVTSRSFGQTGS S SHQKIARDLMT DEVG MKR E  
Sbjct: 360 LGGNDETFKFGSGLLQKQTVDSVTSRSFGQTGGSSSHQKIARDLMTDPVQNMGRHE 419

Query: 539 CLVRIAGVPIERTKYPFLKHKHWKLADKETDDRWW 575  
CLVRIA +PVE++KKY KH +WK LA++ETD+R W  
Sbjct: 420 CLVRIANMPVFKSKYNTKHPNWKYLANQETDERRW 456

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8929> and protein <SEQ ID 8930> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 5  
McG: Discrim Score: 5.53  
GVH: Signal Score (-7.5): -0.78  
Possible site: 34  
>>> Seems to have a cleavable N-term signal seq.  
ALOM program count: 1 value: -7.38 threshold: 0.0  
INTEGRAL Likelihood = -7.38 Transmembrane 66 - 82 ( 62 - 85)  
PERIPHERAL Likelihood = 1.75 338  
modified ALOM score: 1.98

\*\*\* Reasoning Step: 3  
----- Final Results -----  
bacterial membrane --- Certainty=0.3951(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

33.9/50.9% over 419aa

Lactococcus lactis

GP[3582206] trsK protein (traK) Insert characterized

PIR[T43089] transfer complex protein TrsK - plasmid pMRC01 Insert characterized

CRF00383(715 - 2004 of 2415)



-2143-

GP|3582206|gb|AAC56002.1||AE001272(23 - 442 of 530) trsK protein (traK) (Lactococcus lactis)PIR|T43089|T43089 transfer complex prote in TrsK - Lactococcus lactis plasmid pMRC01

```
%Match = 10.1
```

```
%Identity = 3
```

Matches = 141 Mismatches = 193 Cor

PROLOGUE = 29 x PROLOGUE = 299 CORRELATION COEFFICIENT = 76

519            549            579            609            639            669            699            729  
SFLAFILGLVMMTLVILYIVSTGQKYVREGEEYGSARFGTSKKEKNFYSKNPFNDTILARDVRIZTLEKKKQPDNRNKLIL  
| : | : | :  
MNGTILGLVDNKLIIYQDNITKPNRIVM  
10            20

|     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|
| 759 | 789 | 816 | 846 | 876 | 906 | 936 | 966 |
|-----|-----|-----|-----|-----|-----|-----|-----|

VIGGSGAGKTERFKKPNLLOLNCSTIVV-DRKDHLARKTGLFLNGYGVKVI-DLVMTNSDGENPFRYVETENDINRMT

[illegible]

**UUCGCGVHTGGCIGTAT TATTACCTGTTCCKRI YKTAQTKI AGCVTHHIANANMAHSDDVNRPDVTERTFOAECLA**

VIGSGSYKIQSVVITNLFNENFSIVVTDPRGELYEKTAGIKLWQGIKVHVNVNFANMAHSDRINPFDTIERDIQAESVA  
40 50 60 70 80 90 100

40            50            60            70            80            90            100

|     |      |      |      |      |      |      |      |
|-----|------|------|------|------|------|------|------|
| 996 | 1026 | 1056 | 1086 | 1116 | 1146 | 1176 | 1194 |
|-----|------|------|------|------|------|------|------|

TVYFNNTKGNGSRSDPFWDEASMTLVRAIASYLVDLYNPPGSSKQEQEARRKGRYPAFSEIGKLIKLLSKGD---NQD

TKIVOSENAEGKK--DWWFSTOROLLKALILFVM-----KERSPEQRNLGVINVLQTFDSEPINKD

|  |     |     |     |     |     |
|--|-----|-----|-----|-----|-----|
|  | 120 | 130 | 140 | 150 | 160 |
|--|-----|-----|-----|-----|-----|

200 200 200 200 200

|      |      |      |      |      |      |      |      |
|------|------|------|------|------|------|------|------|
| 1221 | 1251 | 1281 | 1311 | 1341 | 1371 | 1401 | 1431 |
|------|------|------|------|------|------|------|------|

1221 1251 1281 1311 1341 1371 1401 1431

R-SILEVLFPEDYAKKYGHENFIMRNWADFQNYLRLTDSVIAVFIARFALFNISVIDETQRDINDLRKRWGTQRIMVILV

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 104

ENSDLDNLF--LALKITHPARIAYELG-FKKAKGDMKASIISLLATISKFTDEEVSNFTSISDFHLQDIGRKKIVLYVI

|     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|
| 180 | 190 | 200 | 210 | 220 | 230 | 240 |
|-----|-----|-----|-----|-----|-----|-----|

1461      1491      1521      1551      1581      1611      1641      1671

IPDNDTTFRFLSALFFSTVFSTLTROADVDFKGOLPIHVRSYLDEFANVGEIPDFAEOTSTVRSRNMSLVPILQNIACL

[illegible]

IRPMDNTVRSPTNLFESOMFDELYKLASSN-GAKLPQEVDFILDFVNIGKEPKYEEFLATCRGYGIGVTTICOTLTOLQ

260 270 280 290 300 310 320

260 270 280 290 300 310 320

SEQ ID 5884 (GBS11d) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 151 (lane 6; MW 61kDa) and in Figure 182 (lane 10; MW 61kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 12 (lane 5; MW 91.5kDa).

### Example 1896

A DNA sequence (GBSx2004) was identified in *S. galactiae* <SEQ ID 5885> which encodes the amino acid sequence <SEQ ID 5886>. Analysis of this protein sequence reveals the following:

Possible site: 50

```
>>> Seems to have no N-terminal signal sequence
```

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4192(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) ≤ sugg>

```

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

60

A related GBS nucleic acid sequence <SEQ ID 9935> which encodes amino acid sequence <SEQ ID 9936> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1897

- 10 A DNA sequence (GBSx2005) was identified in *S.agalactiae* <SEQ ID 5887> which encodes the amino acid sequence <SEQ ID 5888>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3391(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

- 20 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1898

- 25 A DNA sequence (GBSx2006) was identified in *S.agalactiae* <SEQ ID 5889> which encodes the amino acid sequence <SEQ ID 5890>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -10.03 Transmembrane 68 - 84 (64 - 90)
INTEGRAL Likelihood = -7.06 Transmembrane 33 - 49 (27 - 50)
30 INTEGRAL Likelihood = -5.73 Transmembrane 106 - 122 (105 - 123)
INTEGRAL Likelihood = -4.46 Transmembrane 6 - 22 (3 - 24)
INTEGRAL Likelihood = -2.13 Transmembrane 154 - 170 (154 - 170)
INTEGRAL Likelihood = -0.53 Transmembrane 180 - 196 (180 - 196)

35 ----- Final Results -----
bacterial membrane --- Certainty=0.5012(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

- 40 A related GBS nucleic acid sequence <SEQ ID 9937> which encodes amino acid sequence <SEQ ID 9938> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BA11325 GB:D78257 ORF8 [Enterococcus faecalis]
Identities = 35/102 (34%), Positives = 57/102 (55%), Gaps = 4/102 (3%)

45 Query: 90 TRNQAVLVGVGKQVPPPIFLFL-VNASILEIIVYRQLLWEKLTFFPEQIGVTSFLFV 146
T N + L + + + V P + + + E L + A I + E I V + R + L I + S E L F
Sbjct: 7 TANDSTILKLFSGVSPVLVLLLGTAIPIMIRIVFPGGIIGYLVENNALATLISSEFLFG 66
```

Query: 147 LSEGPNQLGSLWLIYSCLGLTLAVVRLKT-DCMTAIALHLLWN 187  
+ HGP S+ +Y +G+ L+V ET D +I++H L N  
Subject: 67 ITHGPTNFISEGMYYFMGII LSVSYKYTKDLRVSIISIHPLNN 108

5

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8931> and protein <SEQ ID 8932> were also identified. Analysis of this protein sequence reveals the following:

```

10 Lipop: Possible site: -1 Crend: 4
 McG: Discrim Score: 9.32
 GvH: Signal Score (-7.5): -5.41
 Possible site: 45
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 6 value: -10.03 threshold: 0.0
15 INTEGRAL Likelihood =-10.03 Transmembrane 68 - 84 (64 - 90)
 INTEGRAL Likelihood =-7.06 Transmembrane 33 - 49 (27 - 50)
 INTEGRAL Likelihood =-5.73 Transmembrane 105 - 122 (105 - 123)
 INTEGRAL Likelihood =-4.46 Transmembrane 6 - 22 (3 - 24)
 INTEGRAL Likelihood =-2.13 Transmembrane 154 - 170 (154 - 170)
20 INTEGRAL Likelihood =-0.53 Transmembrane 180 - 196 (180 - 196)
 PERIPHERAL Likelihood = 1.36 131
 modified ALOM score: 2.51

```

```

25 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5012(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

35 ORF01326(568 - 861 of 1188)  
EGAD[148261][158156(7 - 108 of 120) hypothetical protein {Enterococcus faecalis}  
GP[1402529][dbj][BAA11325.1][D78257 ORF8 {Enterococcus faecalis}  
%Match = 5.9  
%Identity = 34.7 %Similarity = 60.4  
Matches = 35 Mismatches = 37 Conservative Sub.s = 26

[illegible]

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
55 vaccines or diagnostics.

### Example 1899

A DNA sequence (GBSx2007) was identified in *S.galactiae* <SEQ ID 5891> which encodes the amino acid sequence <SEQ ID 5892>. Analysis of this protein sequence reveals the following:

Possible site: 23

-2146-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5           bacterial cytoplasm --- Certainty=0.2490(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9939> which encodes amino acid sequence <SEQ ID 9940> was also identified.

10 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1900

15 A DNA sequence (GBSx2008) was identified in *S.agalactiae* <SEQ ID 5893> which encodes the amino acid sequence <SEQ ID 5894>. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

20       ----- Final Results -----

          bacterial cytoplasm --- Certainty=0.5298(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC98423 GB:L29323 unknown [Streptococcus pneumoniae]

Identities = 68/126 (53%), Positives = 88/126 (68%)

30       Query: 1   MNLHKKSIILDCLEERINQAEINQLQKILSLPNFDCDFEVTEDDYHKEMDPLFYE 60  
           M   L+K+SIILDC ELE +H AE QL ++I +PN+ C+FEVTF DDVHK+ N PLFYE

          Sbjct: 1   MKALNKESILDCLELTELHDAEIKQLDEQIFLMPNYPCEFEVTFLEDYHKGHNPLFYE 60

          Query: 61   SNLHQISDFMETRIDKNGVDITLLTKNHLAFRARGENYSARGKEGILITLTVTKCPGEGR 120  
           S L I +F+E+DIKNG D + +L F +G+ Y A GKEGILIT VIVK F E +

35       Sbjct: 61   SYLQINIMEFLESQDIKNGADAFVDDHQNLFVFLYQGQYRARGKEGILITVTVKAFDEDK 120

          Query: 121 MPIDMS 126

          PI+ +

40       Sbjct: 121 KPINF A 126

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1901

45 A DNA sequence (GBSx2009) was identified in *S.agalactiae* <SEQ ID 5895> which encodes the amino acid sequence <SEQ ID 5896>. This protein is predicted to be methyl transferase. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

50       ----- Final Results -----

-2147-

bacterial cytoplasm --- Certainty=0.1209 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC90421 GB:L29323 methyl transferase [Streptococcus pneumoniae]

Identities = 323/449 (71%), Positives = 389/449 (85%), Gaps = 3/449 (0%)

10 Query: 1 MKFLDLFAGIGGFRGLMESQGEKCLGFCEDIKPARTSYKAMPNTEGEIEYHDIKVEITHD 60  
 M+P+DLF+GIGGFRGLMES GH+G+GFCEDIKPAR SYK++F TEGEIE+HDI++V+D +  
 Sbjct: 1 MRFLDLFAGIGGFRGLMESVGHGICGFCEDIKPARSYKSI+QTEGEIEFHIDIDVDSDDE 60

15 Query: 61 FRQFRGQVDIIICGGFPCQAFSLAGRRIGFEDTRGTLFFETARAQKIQPRFLENVKGIL 120  
 F++ RG+VD+ICGGFPCQAFS+AGRRIGFEDTRGTLFFETARAQKIQPRFLENVKGIL  
 Sbjct: 61 FKKLRGKQVDVICGGFPCQAFSLAGRRIGFEDTRGTLFFETARAQKIQPRFLENVKGIL 120

20 Query: 121 LNHDEGRFTATILSTLDELGYDVGWQLNSKDFQVQNRERFIIIGHSRRYSRIFFLR 180  
 LNH+GRFT TIL+TLDELG+DVGWQ+LNSKDF VQNRERFIIIGHSR+ +R FF R  
 Sbjct: 121 LNH+GRFTTITL+TLDELGFDVGWQLNSKDFQVQNRERFIIIGHSRKRGTLPFPFR 180

25 Query: 181 RED---SFAHLERLGNINPFSKHGLNGEYVLTSLAPTLTRGKGEGAKIAIPVLPDRLEK 237  
 RE +P L+ LGN+NFSK G++G+VY + GLAPTL RKGEG KIAIP +TPDR+K  
 Sbjct: 181 REQQAINPETLKLILGNLNFSGSGMSKGVYSEGLAPTLVRGKGEGKIAIPCMTPDRLEK 240

30 Query: 238 RQGRFRFDKNDQPMFTLTSQDGHGVVAGNLFSDQTRVDFISGLSPTLTITMQSGGKV 297  
 RQ+GRFRFDKNO+PMFTL +QD+HG+VV G+LPTSF +TGRV+ GLSPTLTITMQSGGK+  
 Sbjct: 241 RQGRFRFDKNDQPMFTLNTQDGHGIVVVGDLPTSFKETGRVYSGEGLSPTLTITMQSGGKI 300

35 Query: 298 FKILLREELPFLKIKKATKGVAKATLGSVNLAYPDSTKRRGVGKISNTLTTSDNGS 357  
 FKIL+ E + FLK++EATK GYA+A +GDS+NL P S RGRGVGKI+NTLTS MG  
 Sbjct: 301 FKIL+PEPIQFLKVRKATKGVAKAIGDSINLERPSSQRRGRGVGKIANTLTTSGQNG 360

40 Query: 358 VVVAALSYRQDKWYEVTVIGLEGLYRLRIRRLTPRECFLQGFQFWAYERASSVSSKQ 417  
 VVVA+ E + Y+V G++++G+ YRLRIR+TP+ECFLQGFQFWA+E A VSS SQ  
 Sbjct: 361 VVVA+SYRQDKWYEVTVIGLEGLYRLRIRRLTPRECFLQGFQFWAFAARAKVSSSQ 420

Query: 418 LYKQAGNSVTVPVIAIAIRFRFRTEEEK 446  
 LYKQAGNSVTVPVIAIA++ + EE+++  
 Sbjct: 421 LYKQAGNSVTVPVIAIAAKLKEVEKDE 449

A related DNA sequence was identified in *Sp. pyogenes* <SEQ ID 2435> which encodes the amino acid sequence <SEQ ID 2436>. Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1725 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 60/75 (80%), Positives = 69/75 (92%)

55 Query: 1 MKFLDLFAGIGGFRGLMESQGHKCLGFCEDIKPARTSYKAMPNTEGEIEYHDIKVEITHD 60  
 MKFLDLFAGIGGFRGL+ Q H+C.GFCEDIKPAR SYK++ TEGEIE+HDI++VID D  
 Sbjct: 4 MKFLDLFAGIGGFRGLINOCHECIGFCEDIKPARQSYKAIY+TEGEIEFHIDIROVDQD 63

Query: 61 FRQFRGQVDIIICGGF 75  
 FRQ RGQVDIIICGGF  
 Sbjct: 64 FRQLRGQVDIIICGGF 78

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2148-

**Example 1902**

A DNA sequence (GBSx2010) was identified in *S. agalactiae* <SEQ ID 5897> which encodes the amino acid sequence <SEQ ID 5898>. Analysis of this protein sequence reveals the following:

```

Possible site: 16
5 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -9.71 Transmembrane 8 - 24 (3 - 30)

----- Final Results -----
10 bacterial membrane --- Certainty=0.4885(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9941> which encodes amino acid sequence <SEQ ID 9942> was also identified.

15 The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 5899> which encodes the amino acid sequence <SEQ ID 5900>. Analysis of this protein sequence reveals the following:

```

Possible site: 28
20 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.81 Transmembrane 20 - 36 (19 - 36)

----- Final Results -----
 bacterial membrane --- Certainty=0.1723(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 16/33 (48%), Positives = 26/33 (78%)

30 Query: 1 MNRGMIWVILGGIYLISIIILLIVEIRAPEMDDK 33
 ++KM NW+L G++ + I LI+E+I APEM+D+
 Sbjct: 12 VSGMFVWILLGGVGLRTIKLIIIVITAPEMEDY 44

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1903**

A DNA sequence (GBSx2011) was identified in *S. agalactiae* <SEQ ID 5901> which encodes the amino acid sequence <SEQ ID 5902>. This protein is predicted to be ifn-response binding factor 1 (irebf-1). Analysis of this protein sequence reveals the following:

```

40 Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
45 bacterial cytoplasm --- Certainty=0.4771(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

50 >GP:ADD41248 GB:AF106927 unknown [Streptococcus suis]
 Identities = 258/272 (94%), Positives = 266/272 (96%)

 Query: 1 MKRITANQYQTSERYKLPKILFESERYKDKMLEVKVAYAVLKDRLELSLSEGWIDEIGA 60
 MKRITANQYQTSERYKLPKILFESERYKDKMLEVKVAYAVLKDRLELSLSEGWIDEIGA

```

-2149-

Sbjct: 1 MKRITANQYQTSERYTKPLKILFESERYKDMKLEVKVAYAVLKDRLELSLSKGWIDEGA 60  
 Query: 61 IYLYTNSNLMALLGCSKSLSTIKKTLREYGLIDEVQSSSERGRMANKIYLSELESH 120  
 IYLYTNSNLMALLGCSKSLSTIKKTLREYGLIDEVQSSSE+GRMANKIYLSELESH  
 Sbjct: 61 IYLYTNSNLMALLGCSKSLSTIKKTLREYGLIDEVQSSSERGRMANKIYLSELESH 120  
 Query: 121 TPVLHTDQASVKKTLGSSQRKTGPGVLYSAPSETGSETKYSSETGSDLVMDDEERQVLD 180  
 TPVLHTDQASVKKTLG SQRKTGPGV L SAPSETGSETKYSSET+GSD +++DEERQ VD  
 Sbjct: 121 TPVLHTDQASVKKTLGSSQRKTGPGVLYSAPSETGSETKYSSETGSDFLIDEERQVLD 180  
 Query: 181 EKKEENFTSKVDGVTIKYDRDIWGLVEDQLRQTGLSQSASDYAMIYFSDRYQYALEQMR 240  
 EK+ENFTSKVDGVT+YDRDIWGLVEDQLRQTGLSQSASDYAMIYFSDRYQYALE MRF  
 Sbjct: 181 EKKEENFTSKVDGVTIKYDRDIWGLVEDQLRQTGLSQSASDYAMIYFSDRYQYALEHMR 240  
 Query: 241 ARSAEVIAETVFNGLSEWTKQLRRQVEVGGE 272  
 ARSAEVIAETVFNGLSEWTKQLRRQVEVG+  
 Sbjct: 241 ARSAEVIAETVFNGLSEWTKQLRRQVEVG 272

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5903> which encodes the amino acid  
 sequence <SEQ ID 5904>. Analysis of this protein sequence reveals the following:

Possible site: 55  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.5248 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 84/122 (68%), Positives = 99/122 (80%), Gaps = 2/122 (1%)  
 Query: 145 VLYSAPSETGSETKYSSETGSDLVMDDEERQVLD--EKKEENFTSKVDGVTIKYDRDI 202  
 VL SAPSETE SET+ SET+ S+LV++DEER+ +K E +FT +VD VTKYD+DYI  
 Sbjct: 1 VLYSAPSETEKSETGSETKESNLVIEDEERKECTSVKTKGHPTRQVDQVTKYDKDI 60  
 Query: 203 WGLVHDQLRQTGLSQSASDYAMIYFSDRYQYALEQMRFARSAEVIAETVFNGLSEWTKQ 262  
 W LVH QLR+ GLSQ+ASD M YF +RY YALE +RFAR+AE IAEYVFNGLSEWTKQ  
 Sbjct: 61 WSLVHSQRLREGGLSQASDLVMSYFEERYAYALEHIFPARTABALAEYVFNGLSEWTKQ 120  
 Query: 263 LR 264  
 LR  
 Sbjct: 121 LR 122

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 vaccines or diagnostics.

#### Example 1904

A DNA sequence (GBSx2012) was identified in *S.agalactiae* <SEQ ID 5905> which encodes the amino  
 acid sequence <SEQ ID 5906>. Analysis of this protein sequence reveals the following:

Possible site: 17  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4191 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9375> which encodes amino acid sequence <SEQ ID 9376>  
 was also identified.

-2150-

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### 5 Example 1905

A DNA sequence (GBSx2013) was identified in *S.agalactiae* <SEQ ID 5907> which encodes the amino acid sequence <SEQ ID 5908>. Analysis of this protein sequence reveals the following:

```
10 Possible site: 15
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3723(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
15 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### 20 Example 1906

A DNA sequence (GBSx2014) was identified in *S.agalactiae* <SEQ ID 5909> which encodes the amino acid sequence <SEQ ID 5910>. Analysis of this protein sequence reveals the following:

```
25 Possible site: 44
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3053(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
30 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### 35 Example 1907

A DNA sequence (GBSx2015) was identified in *S.agalactiae* <SEQ ID 5911> which encodes the amino acid sequence <SEQ ID 5912>. This protein is predicted to be 50S ribosomal protein L7/12 (rplL). Analysis of this protein sequence reveals the following:

```
40 Possible site: 56
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1034(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```



-2151-

A related GBS nucleic acid sequence <SEQ ID 9943> which encodes amino acid sequence <SEQ ID 9944> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:CAB11881 GB:Z99104 ribosomal protein L12 (BL9) [Bacillus subtilis]  
Identities = 83/123 (67%), Positives = 95/123 (76%), Gaps = 2/123 (1%)

Query: 6 MALNIENIIAEIKKATILELNDLVKAIIEEFGVTAAAPVAAA--ANGGEAAAKDSFDVE 63  
MALNIE IIA +KEAT+LELNDLVKAIIEEFGVTAAAPV A AA G AA + FD+  
10 Sbjct: 1 MALNIEELIASVKATVLELNDLVKAIIEEFGVTAAAPVAVAGGAAGGAABQSEFDLI 60

Query: 64 LTAAGGKKVGVVIRVREITGSLKEAKAIVDMAFSVIKEGASEARNETKKELEAGASV 123  
L AG +K+ VIKVREITG GLKEAK +VDN P +KEG ++ EA E+K KLE GASV  
15 Sbjct: 61 LAGAGSKQIKIRVIRVREITGSLKEAKELVDNTPKPLKEGLAKEAEELKAKLEEVGASV 120

Query: 124 TLK 126  
+K  
Sbjct: 121 EVK 123

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5913> which encodes the amino acid sequence <SEQ ID 5914>. Analysis of this protein sequence reveals the following:

Possible site: 51  
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.1164 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30 An alignment of the GAS and GBS proteins is shown below.

Identities = 104/126 (82%), Positives = 113/126 (89%)

Query: 1 MEETVMAINENIIAEIKKATILELNDLVKAIIEEFGVTAAAPVAAAAAGGEAAAKDSF 60  
+EITVMAINENIIAEIKK+ILELNDLVKAIIEEFGVTAAAPVAAAAAG AAKDSF  
35 Sbjct: 1 LEEITVMAINENIIAEIKKATILELNDLVKAIIEEFGVTAAAPVAAAAAGGEAAAKDSF 60

Query: 61 DVELTAAAGKKVGVVIRVREITGSLKEAKAIVDMAFSVIKEGASEARNETKKELEAG 120  
DVELT+AGEKKVGVVIRVREITG GLKEAK +VD AP+ +KEG +AEA ETK KLE AG  
40 Sbjct: 61 DVELTSAGKKVGVVIRVREITGSLKEAKGLVDGARNVKKGVAAABEETKAKLEAG 120

Query: 121 ASVTLK 126  
A++TLK  
Sbjct: 121 ATTTLK 126

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1908

A DNA sequence (GBSx2017) was identified in *S.agalactiae* <SEQ ID 5915> which encodes the amino acid sequence <SEQ ID 5916>. This protein is predicted to be ribosomal protein L10 (rplJ). Analysis of this  
50 protein sequence reveals the following:

Possible site: 37  
>>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.1251 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

-2152-

bacterial outside --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the GENPEPT database.

5 >GP: CAB11880 GB: Z99104 ribosomal protein L10 (BL5) [Bacillus subtilis]  
 Identities = 96/164 (58%), Positives = 125/164 (75%), Gaps = 1/164 (0%)

Query: 14 MSEAIIAKKAQVELIAEKMKAAASIVVDSRGLTVQDNLRRSLRSEDVEFKVKNIS 73  
 MS AI KK VE IA K+K + S ++VD RGL V + T LR+ LRE++VE KV KN++  
 10 Sbjct: 1 MSSAISTKKVU-VEETASKIKSEKSTIIVDTRGLNVSEVTELKRLQLEARNVESKYVNTM 59

Query: 74 LTRAAEKAGLEDLKLKELFVGPSAVAFSNEDVIAPAKVISDFAKDAEALIEKGGSDGKFTS 133  
 RA E+A L L + GP+A+AFS EDV+APAKV++DFAK+ EAIEIK G ++GK ++  
 10 Sbjct: 60 TRRAVTRQARLNGLANDFLTGPAIAFSTEDVVAAPAKVINDFAKNHEALIEKAGVIBKGVST 119

Query: 134 VEEINALAKLPNEQMLSMMLSVLQAPVENVAYAVKAAEKDEE 177  
 VER+ ALA+LP +EG+LSMMLSVL+APVEN+ A KAAE+ EE  
 15 Sbjct: 120 VEEVKALAEPLPREGLMLSMMLSVLQAPVENVLAAKAAVQKEE 163

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5917> which encodes the amino acid sequence <SEQ ID 5918>. Analysis of this protein sequence reveals the following:

Possible site: 56  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -5.47 Transmembrane 7 - 23 ( 5 - 24)

25 ----- Final Results -----  
 bacterial membrane --- Certainty=0.3187 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

30 An alignment of the GAS and GBS proteins is shown below.

Identities = 149/176 (84%), Positives = 162/176 (91%)

Query: 4 SQKIKTEVKLMSEAITAKKAEQVELIAEKMKAAASIVVDSRGLTVQDNLRRSLRSED 63  
 S KIKTEVKLMSEAITAKKAEQVELIAEKMKAAASIV+VDSRGLTV+QDT LRRSLRSE  
 35 Sbjct: 23 SPKIKTEVKLMSEAITAKKAEQVELIAEKMKAAASIVVDSRGLTVQDNLRRSLRSE 82

Query: 64 VEFKVIKNSILTRAAEKAGLEDLKLKELFVGPSAVAFSNEDVIAPAKVISDFAKDAEALIEK 123  
 VEFKVIKNSILTRAAEKAGL++LK++FVGPSAVAFSNEDVIAPAKVI+DF K A+ALEIK  
 40 Sbjct: 83 VEFKVIKNSILTRAAEKAGLEDLKLKELFVGPSAVAFSNEDVIAPAKVINDPTKTADALEIK 142

Query: 124 GGSVDGKFTSVEEINALAKLPNEQMLSMMLSVLQAPVENVAYAVKAAEKDEEVA 179  
 GG+++G +S EEI ALA LPN+EGMLSMMLSVLQAPVENVAYAVKAAE E A  
 Sbjct: 143 GGAIEGAVSSKEETQALATLPNRECMMLSMMLSVLQAPVENVAYAVKAAENKEGA 198

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1909

A DNA sequence (GBSx2018) was identified in *S.agalactiae* <SEQ ID 5919> which encodes the amino acid sequence <SEQ ID 5920>. Analysis of this protein sequence reveals the following:

50 Possible site: 40  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -7.22 Transmembrane 125 - 141 ( 106 - 143)  
 INTEGRAL Likelihood = -1.91 Transmembrane 108 - 124 ( 106 - 124)

55 ----- Final Results -----  
 bacterial membrane --- Certainty=0.3688 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

-2153-

A related GBS nucleic acid sequence <SEQ ID 10931> which encodes amino acid sequence <SEQ ID 10932> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1910

A DNA sequence (GBSx2019) was identified in *S.agalactiae* <SEQ ID 5921> which encodes the amino acid sequence <SEQ ID 5922>. This protein is predicted to be Clp-like ATP-dependent protease binding subunit (clpC). Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
15 bacterial cytoplasm --- Certainty=0.3483 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
20 >GP:AAA68910 GB:L34677 Clp-like ATP-dependent protease binding
 subunit [Bos taurus]
 Identities = 437/589 (74%), Positives = 514/589 (87%), Gaps = 5/589 (0%)

25 Query: 10 DFPGN-MDDIFNSLMGRMGVNSNENKRYLINGREVTPPEPFSQRCQTEKLPQGLAQNTPT 68
 Sbjct: 2 DPFN MDD+FN LMG M G NSEN+RYLINGREVTPPEB++ +RGTGKLGNG Q
 DPFNMDLDFNLQMGVNSNENKRYLINGREVTPPEYAFRGTGKLPQGVTDQ-ATK 60

Query: 69 TNQVSADSVLTKLGTNLATQARQHLLDFVIGRNKEIQSTAILARNTQINPVLVGDAGVVG 128
 Sbjct: 61 TQ Q DS+L KLG NLT +A++ LDFVIGRNKEIQSTAIL+RRTKNNPVLVGDAGVVG 120
 TKQFQDSMTAKLRNLTQEAKEGKLDVIGRNKEIQSTAILSRRTKNNPVLVGDAGVVG 120

35 Query: 129 KTAIVGSLAQAIINGDVPAAILNKKEIISIDISLQAGTQVRGSPFENQNTIKVKEFGN 188
 Sbjct: 121 KTAIVGSLAQAI VAGDVPAAILNKQIISIDISLQAGTQVRGSPFENMQKLIDEVKKDGN 180

Query: 189 IILFDEIHQILGAGSTGDSGSKGLADILKPAISRGELTVIGATTDQDEYRNTILKDAAL 248
 Sbjct: 181 IILFDEIHQILGAGSTGDSGSKGLADILKPAISRGELTVIGATTDQDEYRNTILKDAAL 240

40 Query: 249 ARRFNFKVKNAPSAGQTFNIIKLRNLYRQHNVLPDSVLKAVVLSIQYIPQSLPDK 308
 Sbjct: 241 SRRFNQVTYNAPSKESTFKILQGLRGLYKSHHNVSLPDEVLAQAIDSVQYIPQSLPDK 300

Query: 309 AIDLIDNTAAHLAQAHPVTLKSLKKTABQRDKQKAVATDQFERALAKVKTIRKLQNG 368
 Sbjct: 301 AIDLIDNTAAHLA+HPV D K++E+HI + KQ++AV ED++ A + K ++ +LQ+Q 360

45 Query: 369 INHTEQKQVTTATINDIAMSITERTLCVPSVNGASDIERLKLGNRLKGVIGQNDAYEA 428
 Sbjct: 361 LKTHSESEKVVATPSDVAAVSRWGTGIPSVKMGASDIERLKLGNRLKGVIGQNDAYEA 420

Query: 429 VRAIRRNKAGFDGCRNPISGFLPVGPTGVGKTELAKQLAFMPGSKDAIVRLDMSEYID 488
 Sbjct: 421 VRAIRRNKAGFDGCRNPISGFLPVGPTGVGKTELAKQLALNMGSTNDIIRLDMSEYID 480

55 Query: 489 RTAVSKLIGTAGVGYDDNSNTLTETIRRNPSYIVLDEIEKADQPVITLLQLVDQGR 548
 RTAVSKLIG TAGVGYDDNSNTLTET+RR+PYSIVLLDEIEKA+QPVITLLQLVDQGR
```

-2154-

Sbjct: 481 RTAVSKLIGTTAGYVGYDDNSNTLTERKVRHPYSIVLDEIEKANQVITILLQLVLDGR 540

Query: 549 LTDGQNTINFRONTVIIATSNAGFGNKAFTGSDKDKLIMERISPYPRP 597

LTDGQNT++FRONT+IIATSNAGF++A G+ D K+M+++ PYPRP

5 Sbjct: 541 LTDGQNTVDKNTIIATSNAGFSDAVAGE---DAKMDKLQPYPR 586

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5923> which encodes the amino acid sequence <SEQ ID 5924>. Analysis of this protein sequence reveals the following:

Possible site: 22

10 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.2718(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 551/597 (79%), Positives = 616/697 (88%), Gaps = 3/697 (0%)

20 Query: 5 NFYNRDPFGNDDIFNSLMGNMGGYNSENKRYLINGREVTPEEFQYRQTGKLPQQLNN 64  
+F +DFF NMDDIFN LM NMGGY SEN RYL+NGRE+TPEEF YRQTG+LP  
Sbjct: 3 HFSGIDPFVNMDDIFNQLMANMGGYRSKNFRYLNGREITPREFCHYRQTGQVPATTKA 62

25 Query: 65 QNTPTNQVSADSVLTKLGTNLTDQARQHLLDFVIGRNKEIQETAEILARTRKNNPVLVGD 124  
N+ ADSVLT+LGTNLIT +ARQ LDFVIGRNKEIQ+TAEILARTRKNNPVLVGD  
Sbjct: 63 TNSQMLTPKADSVLTQLGTNLTDQARQHLLDFVIGRNKEIQETAEILARTRKNNPVLVGD 122

30 Query: 125 AGVGKTAIVIBLQAQAIINGDVPAAIKKEIISIDISSLEAGTQYRGSFEENIQLNKEVK 184  
AGVGKTAIVIBLQAQAI+NGDVPAAIKKEI+SIDISSLEAGTQYRGSFEE IQN+I+EVK  
Sbjct: 123 AGVGKTAIVIBLQAQAINVDVPAAIKKEIIVSIDISSLEAGTQYRGSFEEIQLNKEVK 182

35 Query: 185 ETGNILFFDRIHQILGAGSTGGDSGSKGLADILKPAISRGELTVIGATTQDEYRNTILK 244  
E GNILFFDRIHQI+GAG+T DSGSKGLADILKPAISRGELT+IGATTQDEYRNTILK  
Sbjct: 183 EAGNILFFDRIHQIVGAGATSSDSGSKGLADILKPAISRGELTILGATTQDEYRNTILK 242

40 Query: 245 NAALARRPFNEVKVNAPSAQOTFNILMGIIRNLVQHHNVLPDSVLKAAVDLSIQYIPQRS 304  
NAALARRPFNEVKVNAPSA+DTP+ILMGIIRNLVQHH++ LPD+VLKAAVD SIQYIPQRS  
Sbjct: 243 NAALARRPFNEVKVNAPSAEOTFHILMGIIRNLVQHHHILPDVNLKAAVDYSIQYIPQRS 302

45 Query: 305 LPDKAIDLIMTAHIAAQAHPVTDLKSLEKIEAQKQKAVANTDFFEALKVKTRIE 364  
LPDKAIDL+IMTAHIAAQAHPVTDL+L E ELA+Q++ QEKAV EDPE+AL KTRIE  
Sbjct: 303 LPDKAIDLIMTAHIAAQAHPVTDLKTLETIAKQKQKAVAKEDFEKALAKTRIE 362

50 Query: 365 LQNQIDNHTEGQKVTATINDIAMSIERITGVPSVNMAGSDIERLKGELNRLLKGVQIQND 424  
LQ QI+ H + Q VTAT+NDIA S+ERLTO+PVSNMG +D+ERLK + +RLK VIGQ++  
Sbjct: 363 LQKQIEQHNSQNTATVNDIARSVERLTGIPVSNMGNDLERLKGISRLKSHVIGQDE 422

Query: 425 AVEAVARAIRRNKAGFDDGNRPISGFLFVGPTGVGKTEIAKQLADFMFGSKDAIVRLDMS 484  
AV AVARAIRRNKAGFDDG RPISGFLFVGPTGVGKTEIAKQLA D+FGSKDAI+RLDMS  
50 Sbjct: 423 AVEAVARAIRRNKAGFDDGKRPISGFLFVGPTGVGKTEIAKQLALDLMFGSKDAIRLDMS 482

Query: 485 EYNDRTAVSKLIGTAGYVGYDDNSNTLTERIRNRPYSIVLDEIEKADPOVITILLQVL 544  
EYNDRTAVSKLIG TAGYVGYDDN+NTLTER+RNPY+IVLDEIEKADPO+ITILLQVL  
50 Sbjct: 483 EYNDRTAVSKLIGTAGYVGYDDNNNTLTERVRNPYIAIVLDEIEKADPOITILLQVL 542

55 Query: 545 DDGRITDQGNNTINFRONTVIIATSNAGFGNKAFTGSDKDKLIMERISPYPRPEPLNRPN 604  
DDGRITDQGNNTINFRONTVIIATSNAGFG + + IM+RI+PYPRPEPLNRPN  
Sbjct: 543 DDGRITDQGNNTINFRONTVIIATSNAGFGGQ---DTETSSNMRIAPYPRPEPLNRPN 599

60 Query: 605 GVIEPSHLSKDDLSIVDMLDEVNQITGKKGIDLVDSNVKSHLIRLGYDEAMGVRPLR 664  
+I+P+HL K L RIVDML EVNQIT KGI L + ++ K+HLI+IGY+ AMG RPLR  
Sbjct: 600 SIIFKPHLQKSLSEIVDMLAEVNTQITAKKGISLITTDKAGHLIDIGYHAMGARPLR 659

Query: 665 RVIBQIEIRDRITTYLDETDVKHLKANLQDQGVISE 701  
R+IBQIEIRDRITTYLDE+VK L+A L++GQ+VI +  
65

-2155-

Sbjct: 660 RIIEQRIRDRITDYLLDHPFVKLQAILKEGQLVIRQ 696

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 5 Example 1911

A DNA sequence (GBSx2020) was identified in *S. agalactiae* <SEQ ID 5925> which encodes the amino acid sequence <SEQ ID 5926>. Analysis of this protein sequence reveals the following:

Possible site: 20  
 >>> Seems to have an uncleavable N-term signal seq  
 10 INTEGRAL Likelihood = -4.78 Transmembrane 8 - 24 ( 7 - 25)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2911 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 15 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9945> which encodes amino acid sequence <SEQ ID 9946> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:AACT73364 GB:AB000134 putative enzyme [Escherichia coli K12]  
 Identities = 142/307 (46%), Positives = 195/307 (63%), Gaps = 6/307 (1%)  
 Query: 39 KELLESKKTLLILNGALGTELSRGCDVSGKLWSAKYLIEDPAAIQTTHEDYIRAGADIVT 98  
 + LL+ + L+L GA+ TELE+RCG+++ LWSAK L+E+P I+ +H DY RAGA  
 25 Sbjct: 8 RALLDKQDILLDDGAVTELEARGCNLADSLWSAKVLVENPELIREVHLLDYRAGAQCQAI 67  
 Query: 99 TSTYQATLQGLAQGVGSQESQETDLRLITVOLAKAAREQWKSLEKEKSERIYPLISGDV 158  
 T++YQAT G A G+ E+Q++ LI +V+LA+ ARE L + ++ + L++G V  
 30 Sbjct: 68 TASTYQATPAGFAARGLDEAQSALIGKSVELARIAREAY---LAENPQAGTL--LVAGSV 122  
 Query: 159 GPYAAFLADGSEYTGLYDIDKQGLKNFHRHRIELLLDEGVLDIALETIPNAQBAEALIEL 218  
 GPY A+LADGSEY G Y + + PHR R+E LLD G D+LA ET+PN E EAL EL  
 35 Sbjct: 123 GPYGAFLADGSEYRGDYHCSVEAPQAQPHRPRVEALLDAGADLLACETLPNPSIEBALIEL 162  
 Query: 219 LAEDFPQVEAYMSFTSQDGTISDGSAAVADLAKAIDVSPQVVALGINCSSPSLIVADFLQA 278  
 L +P+ A+ SFT +D+ +SDG+ +D+ + PQVVALGINC + LQ  
 40 Sbjct: 183 LTA-YPRARAWFSFTLRDSEHLSGDTPLRDVALLAGYPQVVALGINCIALENTITAAALQ 241  
 Query: 279 IABQINKPLVITYPNSGEVYDQASQWQSSPDHSHTLLENTSDWQKLGAQVVGCCRTFPA 338  
 + T PLV YPNSGE YD S++W +H L + WQ GR+++GGSCRT FPA  
 45 Sbjct: 242 LHGLTVLPLVVPYNSGEHYDAVSKTWHHGHCAQLADYLPQWQAAGARLIGGCCRTFPA 301  
 Query: 339 DIADLSA 345  
 DIA L A  
 50 Sbjct: 302 DIAALKA 308

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

50 A related GBS gene <SEQ ID 8933> and protein <SEQ ID 8934> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 5  
 MoG: Discrim Score: 5.48  
 GvH: Signal Score (-7.5): -2.64  
 55 Possible site: 20

-2156-

```
>>> Seems to have an uncleavable N-term signal seq
ALOM program count: 1 value; -4.78 threshold: 0.0
 INTEGRAL Likelihood = -4.78 Transmembrane 8 - 24 (7 - 25)
 PERIPHERAL Likelihood = 2.49 259
modified ALOM score: 1.46
```

\*\*\* Reasoning Step: 3

```

----- Final Results -----
 bacterial membrane --- Certainty=0.2911(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

CRF01312(412 - 1338 of 1644)  
CMNI|NT01EC0303(55 - 357 of 358) conserved hypothetical protein  
%Match = 23.8  
%Identity = 46.6 %Similarity = 64.3  
Matches = 142 Mismatches = 107 Conservative Sub.s = 54

288            318            348            378            408            438            468            498  
LISQSPCS\*FRI\*GLIGIAHNVLGFTSVFHLLFSAIPITYVTRNGDLGRFKELLESKKTLLILGALGTESRGCDVS  
     | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
ANMPVLGHHSIQRELRQAGYRLLRCAMVLSILNPETQRNSNMSPFLRALDKODILLGDGMATELRAGNTLA

-20-                  -40-                  -60-                  -80-

[illegible][illegible][illegible]

1248      1278      1308      1338      1368      1398      1428      1458  
SPDHSHTLENTSDWKLGAQVGGCCRTPADLSAHLK\*VKYLEEG\*GKPDFLQPSRTRKPAWLIPNGFCFYLSMT  
z | : | : || ::||| ||||| ||||| ||  
HGEHCALADYLPQWAAGARLLGGCCRTTADIAALKARS  
320            340            360

SEQ ID 8934 (GBS381) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 6; MW 42kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 4; MW 66.9kDa).

### Example 1912

A DNA sequence (GBSx2021) was identified in *S.galactiae* <SEQ ID 5927> which encodes the amino acid sequence <SEQ ID 5928>. Analysis of this protein sequence reveals the following:

Possible site: 51  
>>> Seems to have no N-terminal signal sequence

```

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2995(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

```

-2157-

bacterial outside --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1913

A DNA sequence (GBSx2022) was identified in *S.agalactiae* <SEQ ID 5929> which encodes the amino acid sequence <SEQ ID 5930>. Analysis of this protein sequence reveals the following:

10 Possible site: 59  
 >>> Seems to have no N-terminal signal sequence

|             |                     |               |                        |
|-------------|---------------------|---------------|------------------------|
| INTEGRAL    | Likelihood = -11.62 | Transmembrane | 176 - 192 ( 168 - 198) |
| INTEGRAL    | Likelihood = -11.57 | Transmembrane | 89 - 105 ( 80 - 111)   |
| INTEGRAL    | Likelihood = -10.03 | Transmembrane | 337 - 353 ( 332 - 359) |
| 15 INTEGRAL | Likelihood = -9.87  | Transmembrane | 292 - 308 ( 285 - 316) |
| INTEGRAL    | Likelihood = -4.51  | Transmembrane | 58 - 74 ( 52 - 78)     |
| INTEGRAL    | Likelihood = -3.88  | Transmembrane | 267 - 283 ( 267 - 286) |
| INTEGRAL    | Likelihood = -3.08  | Transmembrane | 125 - 141 ( 125 - 142) |
| 20 INTEGRAL | Likelihood = -2.13  | Transmembrane | 212 - 228 ( 212 - 228) |

20 ----- Final Results -----  
 bacterial membrane --- Certainty=0.5649 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

25 A related GBS nucleic acid sequence <SEQ ID 9377> which encodes amino acid sequence <SEQ ID 9378> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:CAB12034 GB:Z99105 similar to histidine permease [Bacillus subtilis]  
 Identities = 221/384 (57%), Positives = 291/384 (75%), Gaps = 2/384 (0%)

Query: 2 PVTGSPHTYATKFISSGTGFTVANLYWICHTVALGTEPLGAMLAQRWFENFVAFAPAF 61  
 PVTGSPHTYA KAI ROTGFTVANLYW+ WTVALG+EF A +LMQWPF+ W +++  
 35 Sbjct: 76 PVTGSPHTYAKYIGRTGTFTVANLYWLTWTVALGSEPTAGLLMQRWPFHTSVVMSEAV 135

Query: 62 PALVIFGLNALSVRFFAEASFSSSIKVIATIIPIILGLGAMFOLVSPEQKAILFTHL 121  
 PAL IF LNA SV+PFAR+E +FSSIKV+AI++PI+LG AMFG++ +G A ++  
 40 Sbjct: 136 PALPILFGLNALSVKFFAEASFSSSIKVLATVILFGLGAMFOLVSPEQKAILFTHL 195

Query: 122 TANGA-FPNGIVAVSVMLAVNYAFSGTELIGIAAGETDNPKEAVPRAIKTTIGRLVVF 180  
 TA G FENG V ++ ML+VN+AFSGTELIGIAAGR+ +P + +PAIKTT+ RL +FF  
 45 Sbjct: 196 TAEGGLFNGFVPIHMLSVNFASFSGTELIGIAAGSEVDPKTIPKAIKTTVRLGLGF 255

Query: 181 VLTIVVLASLLPMKKGAVSTAPFVDVDFDMGIPFTADIMNFVILTALISAGNSGLYASSR 240  
 V TI VL+ L+P+++AGV +PFV VFD++G+P+ NDMNFVILTALISAGNSGLYASSR  
 50 Sbjct: 256 VGTIFVLSGLPIQAGVVKSPFVAVFDVGVYAADIMNFVILTALISAGNSGLYASSR 315

Query: 241 MLNSLANEGMLSKSVVKINKGVPMRALLSMAGAVLSLSSSIYAADTVILVLSIAGFA 300  
 MLNSL+ E L + K+ G P AL+ SM G +LSL SS++A DIVY+ LVSIGFA  
 55 Sbjct: 316 MLNSLSKKTALHPTAKPI+SKTTFNALVFSMIGSILSSSVFAPDVTVVVLSIGFA 375

Query: 301 VVVVLAIPVAQINFRKEFLKE-NQLKDLSTYKTPFTVPVLYITILLISIVIGIADSSQ 359  
 VVVVW+ I +Q FRK +++ N++ DL Y+TP P +P +L L S+VGIA+D +Q  
 60 Sbjct: 376 VVVVWGLIASQPMFRKRYIRAGNKVTDIKYRTPLYPVFPVIAAFLCLASVVGIVADPNQ 435

Query: 360 RAGLYPGVDFILPCYIYHKLRYKK 383  
 R LY GVFF+ CY + ++ +K  
 65 Sbjct: 436 RIALYCGVDFILPCYIYVKNR 459

There is also homology to SEQ ID 4070.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### 5 Example 1914

A DNA sequence (GBSx2023) was identified in *S.agalactiae* <SEQ ID 5931> which encodes the amino acid sequence <SEQ ID 5932>. Analysis of this protein sequence reveals the following:

Possible site: 32  
>>> Seems to have no N-terminal signal sequence  
----- Final Results -----  
bacterial cytoplasm --- Certainty=0.2378(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

There is also homology to SEQ ID 5642.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1915

20 A DNA sequence (GBSx2024) was identified in *S.agalactiae* <SEQ ID 5933> which encodes the amino acid sequence <SEQ ID 5934>. Analysis of this protein sequence reveals the following:

Possible site: 25  
>>> Seems to have no N-terminal signal sequence  
----- Final Results -----  
bacterial cytoplasm --- Certainty=0.4935(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1916

35 A DNA sequence (GBSx2025) was identified in *S.agalactiae* <SEQ ID 5935> which encodes the amino acid sequence <SEQ ID 5936>. Analysis of this protein sequence reveals the following:

Possible site: 42  
>>> Seems to have no N-terminal signal sequence  
----- Final Results -----  
bacterial cytoplasm --- Certainty=0.0530(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.



-2159-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1917**

A DNA sequence (GBSx2026) was identified in *S. agalactiae* <SEQ ID 5937> which encodes the amino acid sequence <SEQ ID 5938>. Analysis of this protein sequence reveals the following:

Possible site: 17  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0175 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF63739 GB:AF236863 hypothetical GTP-binding protein  
[Lactococcus lactis]  
Identities = 142/193 (73%), Positives = 165/193 (84%)

Query: 6 LNTNHSILLSAANKSHYPQDDLPEVALAGRSNVGKSSPINTILGRNLARTSSKIGKTQ 65  
+NT+N +I +SAA+K YP++D PE+ALAGRSNVGKSSPINTIL RKN ARTS +FGKTQ  
Sbjct: 3 INTNLTITTSAAASKQYPENDWPEIALAGRSNVGKSSPINTILLRKNFARTSSQFGKTQ 62

Query: 66 LLNFYNIDDKLRFPVDVPGYGYAKVSKTERAKWGMIEEVLTRDNLRAVVSIVDFRHPDS 125  
LLNFYNIDDL-L FVDVPGYGYA+VSK ER KMGWMIIEHL TR+NL+ VVSLVD RH+PS  
Sbjct: 63 LLNFYNIDDLHFPVDVPGYGYAKVSKTERAKWGMIEEVLTRDNLRAVVSIVDFRHPDS 122

Query: 126 ADDIQMYEFLKYYIIPVIVATKADKIPRGKWNKHSSIIKKLNFDPKDKHIVPSSSVRT 185  
DD+ MYEFLKYY IPVI+VATKADK+PRGKWNKHES IKK + FD D FI+PSS D+T  
Sbjct: 123 EDDLAMYEFLKYYIIPVIVATKADKIPRGKWNKHESIIKKAMKFDSTDDFIIPSSDTKT 182

Query: 186 GLDESVDITLSEL 198  
G++E+W IL L  
Sbjct: 183 GIEEAWTALKYL 195

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 5939> which encodes the amino acid sequence <SEQ ID 5940>. Analysis of this protein sequence reveals the following:

Possible site: 18  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0123 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 167/196 (85%), Positives = 183/196 (93%)

Query: 3 EEFINTNHSILLSAANKSHYPQDDLPEVALAGRSNVGKSSPINTILGRNLARTSSKPG 62  
E+ LNTNHSILLSAANKSHYPQDDLPE+ALAGRSNVGKSSPINTILGRNLARTSSKPG  
Sbjct: 4 EGVINTNHSILLSAANKSHYPQDDLPEIALAGRSNVGKSSPINTILGRNLARTSSKPG 63

Query: 63 KTQLLAFFNIDDKLRFPVDVPGYGYAKVSKTERAKWGMIEEVLTRDNLRAVVSIVDFRHP 122  
KTQLLAFF+NTDDKLRFPVDVPGYGYAKVSK+ERAKWGMIEEVL+RDNLR VVSLVD RH  
Sbjct: 64 KTQLLAFFNIDDKLRFPVDVPGYGYAKVSKERAKWGMIEEVL+TRDNLRAVVSIVDFRHP 123

Query: 123 DPSADDIQMYEFLKYYIIPVIVATKADKIPRGKWNKHSSIIKKLNFDPKDKHIVPSSSV 182  
PS +DIQMY+FLKYY+IPVI+VATKADKIPRGKWNKHES +KK LHFDK D FLVPSV  
Sbjct: 124 AFSKSDIQMYDPLKYYIIPVIVATKADKIPRGKWNKHESVKKLNFDPKDKHIVPSSSV 183

-2160-

Query: 183 DRTGLDESMDTILSEL 198  
 +R G+D+SND IL ++  
 Sbjct: 184 ERIGIDSGMDAILQV 199

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1918

- A DNA sequence (GBSx2027) was identified in *S. agalactiae* <SEQ ID 5941> which encodes the amino acid sequence <SEQ ID 5942>. This protein is predicted to be protease ClpX (clpX). Analysis of this protein sequence reveals the following:

Possible site: 50  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.2389 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- A related GBS nucleic acid sequence <SEQ ID 9947> which encodes amino acid sequence <SEQ ID 9948> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

- >GP:AAF63738 GB:AF236863 protease ClpX [Lactococcus lactis]  
 Identities = 305/395 (77%), Positives = 357/395 (90%), Gaps = 1/395 (0%)
- 25 Query: 18 NVYCSFGKSGQDEVKIIAGNGVFIQNECVALSQRIIKKEELAEVLAIAEVPKPKELLE 77  
 N+ CSFGKSGQD+VRK+IAG+ V+IQNEC+ LS I++EEL EE +++ EV PKE+ +  
 Sbjct: 8 NIQCSFGKSGQDDVKMIAGSDVYICNECIELSTRILEELKEEQDSEMLEVKTPEKFD 67
- 30 Query: 78 ILNQYVVGQDRAKRALAVAVYHYKRVSYTSS- DDDVDLQKSNILMIGPTGSGKTFLAQ 136  
 LN+YV+QG+AKRALAVAVYHYKRV++T S +D++LQKSNIL+IGPTGSGKTFLAQ  
 Sbjct: 68 HLNQYVVGQEKAKRALAVAVYHYKRVINFTASKIAEDIELQKSNILLIGPTGSGKTFLAQ 127
- 35 Query: 137 TLAKSLNVFPAIADATSLTEAGYVGEDVENILLKLIQAADINVERAERGIIYVDEIDKIA 195  
 TLAKSLNVFPAIADATSLTEAGYVGEDVENILLKLIQAADINVERAERGIIYVDEIDKIA  
 Sbjct: 128 TLAKSLNVFPAIADATSLTEAGYVGEDVENILLKLIQAADINVERAERGIIYVDEIDKIA 187
- 40 Query: 197 KKGENVISITRDVSGEGVQALLKIIIEGTVASVPPQGRKHPNQEMIQITKNILFIVGGA 256  
 KK ENVISITRDVSGEGVQALLKIIIEGTVASVPPQGRKHPNQEMIQITKNILFIVGGA  
 Sbjct: 188 KKGENVISITRDVSGEGVQALLKIIIEGTVASVPPQGRKHPNQEMIQITKNILFIVGGA 247
- 45 Query: 257 FDGIEDLVKQRLSEKVIQFGPQTSRKIDNASYMQRIISEDIOKQFGLIPRFTIGRLPVVAAL 316  
 FDGIE++VKQRLSEK+IQFG ++K+ D SYMQRII+EDIOKQFGLIPRFTIGRLPVVAAL  
 Sbjct: 248 FDGIEIVKQRLSEKVIQFGVANNKKISDEDSYMQRIIASEDIOKQFGLIPRFTIGRLPVVAAL 307
- 50 Query: 317 ELLTAEDLVRLITRPNALVKQYQVLTLLSDYGVKLEFQDALLAIDAKAIERKTIGARGLRS 376  
 E LT EDL++ILTRP+NAL+KQY+ LL +D VLELF AL+AIA KAIBKRTIGARGLRS  
 Sbjct: 308 ERLTIEDLIQILTRPNALIKQYQVLTLLSDYGVKLEFQDALLAIDAKAIAIERKTIGARGLRS 367
- Query: 377 IIEETMIDIMPEVPSQEDVTKVRITKAABVGDKP 411  
 IIEE M+DIMP+PS E++TKV IT+A V+G +P  
 Sbjct: 368 IIEEVMDIMPEVPSHEETKVITTEAVVDGKARP 402

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 5943> which encodes the amino acid sequence <SEQ ID 5944>. Analysis of this protein sequence reveals the following:

- 55 Possible site: 42  
 >>> Seems to have no N-terminal signal sequence
- Final Results -----

-2161-

bacterial cytoplasm --- Certainty=0.2711 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

# 5 An alignment of the GAS and GBS proteins is shown below.

Identities = 378/409 (92%), Positives = 393/409 (95%), Gaps = 1/409 (0%)

```

Query: 9 MAGNRNNDNMVYCSFOGKSGQDEVKIIAGNGVFIQNECVALSQEIIEBELAREVLADLAE 68
10 Sbjct: 1 MAGSRINDIKVYCSFOGKSGQDDVKIIAGNNVFIQNECVALSQEIIEBELAREVLADITE 60

Query: 69 VPKPKELLEILNQYVVQGDRAKRALAVAVNHYKRVSYTES-SDDVDLQKSNILMITGPT 127
10 Sbjct: 61 VPKPKELLEILNQYVVQGDRAKRALAVAVNHYKRVSYTES-DDVDLQKSNILMITGPT 120

15 Query: 128 GSKCTFLAQTAKSLNVFFAIDATSLTEAGTVGEDVENILKKLIQAADYNVERAERGII 187
 GSKCTFLAQTAKSLNVFFAIDATSLTEAGTVGEDVENILKKLIQAADYNVERAERGII
10 Sbjct: 121 GSKCTFLAQTAKSLNVFFAIDATSLTEAGTVGEDVENILKKLIQAADYNVERAERGII 180

20 Query: 188 YVDEIDKIAKKGENVISITROVSGEGVQALLKIIERTVASVPPQGGRRKHPNQEMIQITK 247
 YVDEIDKIAKKGENVISITROVSGEGVQALLKIIERTVASVPPQGGRRKHPNQEMIQI+TK
10 Sbjct: 181 YVDEIDKIAKKGENVISITROVSGEGVQALLKIIERTVASVPPQGGRRKHPNQEMIQIDTK 240

Query: 248 NILPTVGGAFDGIEDLVKQRLGKVGIFGQTSRKIDDNASYMQEIIISEDQKFGLIPEFI 307
10 Sbjct: 241 NILPTVGGAFDGIIE+VKQRLGKVGIFGQ SRKIDDNASYMQEIIISEDQKFGLIPEFI 300

Query: 308 GRLEPVVALELLTARDLVKRLTEPRNALVKQYQILLSDYGVLEFDQDALLAIDAKAIER 367
10 Sbjct: 301 GRLEPVVALE L DL++ILTEPRNALVKQYQ LLSYDGVLE FD++AL AIA+KAIER 360

Query: 368 KTGARGLRSIIEETMLDIMEFIPQSDVTKVRITTKAAVSGTDKPVLETA 416
 KTGARGLRSIIEETMLDIMEFIPQSDVTKVRITTKAAVEG KPVLETA
10 Sbjct: 361 KTGARGLRSIIEETMLDIMEFIPQSDVTKVRITTKAAVEGSKPVLETA 409

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 1919

40 A DNA sequence (GBSx2028) was identified in *S.agalactiae* <SEQ ID 5945> which encodes the amino acid sequence <SEQ ID 5946>. Analysis of this protein sequence reveals the following:

Possible site: 26  
 >>> Seems to have an uncleavable N-term signal seq

```

----- Final Results -----
45 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

50 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## Example 1920

55 A DNA sequence (GBSx2029) was identified in *S.agalactiae* <SEQ ID 5947> which encodes the amino acid sequence <SEQ ID 5948>. Analysis of this protein sequence reveals the following:

-2162-

Possible site: 36  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 5                   bacterial cytoplasm --- Certainty=0.4029(Affirmative) < succ>  
                  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9949> which encodes amino acid sequence <SEQ ID 9950> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC33872 GB:AF055727 dihydrofolate reductase [Streptococcus pneumoniae]  
Identities = 83/162 (51%), Positives = 118/162 (72%), Gaps = 1/162 (0%)

- 15   Query: 25   MTKQIIIAWADEEDHLIGVNGSLPWRLPKELHFKKETTIMGQALLMRKTFDGMNRRVLPG 84  
              MTK+I+AIWA+DE+ LIG   LW LP EL HFKETT+   A+LMGR TFDGM RR+LP  
      Sbjct: 1   MTKKIIVAWQDEBGLIGKENRLPWHLPAELQHFKETTILNHAILMGRVTFDGMRRLLPEK 60
- 20   Query: 85   RETIILTKEQFQADGVTVLNSVEQVIKWFQEHNKLTIFVGGASIIYKALFYCEAIITK 144  
              RET+ILT++ + + DGW   V+ V+ W+Q+ K L+I+GG   I++AF PY + + I T  
      Sbjct: 61   RETLLITRNPEEKIDGVATFQDVQSVLDWYQDQERNLYIIGGKQIFQAFEPYLDDEVIVTH 120
- Query: 145 VHKGPKGDTYFP-DVNLSEPKVISRDYFEKDEQNAHAFTVY 185  
              +H + +GDTYFP +++LS F+ +S   + KDE+N + FT+ Y  
25       Sbjct: 121 I HARVEGDTYFESLDLSPETVSSKPYAKDEKNPYDFTIQY 162

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5949> which encodes the amino acid sequence <SEQ ID 5950>. Analysis of this protein sequence reveals the following:

Possible site: 51  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 30                   bacterial cytoplasm --- Certainty=0.1214(Affirmative) < succ>  
                  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
35                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 82/160 (51%), Positives = 119/160 (74%).

- 40   Query: 25   MTKQIIIAWADEEDHLIGVNGSLPWRLPKELHFKKETTIMGQALLMRKTFDGMNRRVLPG 84  
              MTK+IIIAWADE   LIG+ G LPW LPKEL HFK+IT+ QA+LMGR TP+GMN + LP  
      Sbjct: 1   MTKEIIIAWADEAGLIGIAGKLPWYLPKELEHFKKTTLHQAIIIMGRVTFEGMNRCKRLQ 60
- 45   Query: 85   RETIILTKEQFQADGVTVLNSVEQVIKWFQEHNKLTIFVGGASIIYKALFYCEAIITK 144  
              R+T+++T++ +Q D V   +S+E+V++W+   +KTL+I+GG + +AF Y + IKT  
      Sbjct: 61   RQTILVMTNRNDYQVDEVLTMTSIEKVLWYHQAQDKTYLIIGKNVLEAFNGYFDRIKTV 120
- Query: 145 VHKGPKGDTYFPDVNLSEPKVISRDYFEKDEQNAHAFTVY 184  
              +H +FKGDTY P+++ S F   S+ ++ +D +N + FTVT  
50       Sbjct: 121 IHRHFKGDTYRPMPLDSHFTYQESQTFYARDAKNPYDFTVT 160

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1921

- 55   A DNA sequence (GBSx2030) was identified in *S.galactiae* <SEQ ID 5951> which encodes the amino acid sequence <SEQ ID 5952>. Analysis of this protein sequence reveals the following:

Possible site: 45

-2163-

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

5           bacterial cytoplasm --- Certainty=0.1577 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAA25221 GB:M33770 thymidylate synthase (EC 2.1.1.45)  
           [Lactococcus lactis]  
           Identities = 215/280 (76%), Positives = 245/280 (86%), Gaps = 2/280 (0%)

Query: 1   MTKADLLFKDNITKIMSBGVFSBQARPRYKNGEMANSKYITGAPAEYDLSKGEPFITTLR 60  
           MT AD +FK NI I+ GVPSS ARP+YK+G+MANSKY+TG+F YDL KGEFPITTLR  
 15 Sbjct: 1   MTYADQVFQKNIQNLINQVPSRNARPKYKDGQANSKYVTGSPVYDILQKGKGFPTTLR 60

Query: 61   PIPIKSAIKEIFWIYQDQTNDLAVLNNDKYGVTYWNWVEGHTGTIGQRYGAVVKKHNIIS 120  
           PIPIKSAIKE+ WIYQDQ+L+VL +KYGV YW +W +G GTIGQRYGA VRK+NII  
 20 Sbjct: 61   PIPIKSAIKELMWIYQDQTSLSVLSEKYGKVGKWSGIGD-GTIGQRYGATVKYKNIIG 119

Query: 121   KLLKQLENDPNWNRNRINSLMDYFAESTEGLLPCAFQTMFDVRRV-NGELYDLATLTQRS 179  
           KLL+ L NPWNRNRN+I+LW YE FESTBGLLPCAFQTMFDVRR +G++YLDATL QRS  
 25 Sbjct: 120   KLLKGLAENPNWNRNRINSLWQYEDFESTBGLLPCAFQTMFDVRRKQDQIYLDATLTQRS 179

Query: 180   NMLVAHHINAMQYVALQMIAKHFGWRVGKFFYFINNLHIYDNQFEQAQLLKRQPSBC 239  
           NMLVAHHINAMQYVALQMIAKHG W+VGKFFYF+NNLHIYDNQFEQA EL+KR SE  
 30 Sbjct: 180   NMLVAHHINAMQYVALQMIAKHGFWKVGKFFYFVNNLHIYDNQFEQAQLLKRITASEK 239

Query: 240   NPKLVNLVDPDGTDFFDIKPDDFALVDYDIKPKQLFDLAI 279  
           P+LVLNLVDPGT+FPDIKP+DF LVDY+P+KPQL+FDLAI  
 30 Sbjct: 240   EPRLVLNLVDPGTNFFDIKPDDFELVDYEPVKPQLFDLAI 279

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5953> which encodes the amino acid sequence <SEQ ID 5954>. Analysis of this protein sequence reveals the following:

35       Possible site: 53  
       >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40           bacterial cytoplasm --- Certainty=0.3131 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 227/279 (81%), Positives = 251/279 (89%)

45 Query: 1   MTKADLLFKDNITKIMSBGVFSBQARPRYKNGEMANSKYITGAPAEYDLSKGEPFITTLR 60  
           MTKAD +FK NI KI++BG SBQARP+YK+G A+SKYITGAPAEYD+KGEFPITTLR  
           Sbjct: 9   MTKADQIFKANIQKINSGSIQBQARPKYKDGRTAHSKYITGAPAEYDLAKGEFPITTLR 68

50 Query: 61   PIPIKSAIKEIFWIYQDQTNDLAVLNNDKYGVTYWNWVEGHTGTIGQRYGAVVKKHNIIS 120  
           PIPIKSAIKE+FWIYQDQ+H L VL KY Y YWN+WEV T TIGQRYGAVVKKH+IIS  
           Sbjct: 69   PIPIKSAIKELFWIYQDQNSLDVLKAKYNVHWNWVEVDQTRTIGQRYGAVVKKHDIIS 128

55 Query: 121   KLLKQLENDPNWNRNRINSLMDYFAESTEGLLPCAFQTMFDVRRV-NGELYDLATLTQRSN 180  
           K+LQQL +NPWNRNRINSLMDYFAEST+GLLPCAFQ MFDVRRV +LYLDA+LTQRSN  
           Sbjct: 129   KLLKQLAENPNWNRNRINSLMDYFAESTKGLLPCAFQTMFDVRRVGBDLYLDA+LTQRSN 188

60 Query: 181   DMLVAHHINAMQYVALQMIAKHFGWRVGKFFYFINNLHIYDNQFQAQLLKRQPSBCN 240  
           D+LVAHHINAMQYVALQMIAKHFGW++GKFFYF+NNLHIYDNQF+QAQLLKRQ  
           Sbjct: 189   DILVAHHINAMQYVALQMIAKHFGWKIGKFFYFVNNLHIYDNQFQAQLLKRQPSVASQ 248

Query: 241   PKVLNLVDPDGTDFFDIKPDDFALVDYDIKPKQLFDLAI 279  
           PKVLNLVDP T+FPDIKPDDF L +YDP+KPQL FDLAI  
           Sbjct: 249   PKVLNLVDPDRTNFFDIKPDDFELQNYDVPVKPQLHFDLAI 287

-2164-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1922

- 5 A DNA sequence (GBSx2031) was identified in *S.agalactiae* <SEQ ID 5955> which encodes the amino acid sequence <SEQ ID 5956>. This protein is predicted to be HMG-CoA synthase. Analysis of this protein sequence reveals the following:

Possible site: 60  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.0816 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5957> which encodes the amino acid sequence <SEQ ID 5958>. Analysis of this protein sequence reveals the following:

Possible site: 58  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.1670 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 260/385 (67%), Positives = 325/385 (83%)

Query: 36 MKIGIDKIGFATSQYVLEMTDLAARQVDPKFKSGLLLSLITPTVEDITVLAASRN 95  
 M IGIDKIGFATSQYVL++ DLA+ARQVDP KFS+GLL++S S+ P+TEDI+TLAASRN  
 30 Sbjct: 14 MTIGIDKIGFATSQYVLEMTDLAARQVDPKFKSGLLLSLITPTVEDITVLAASRN 73

Query: 96 DILGDEKKEITDMVIVATSSIDQSKAASVYVHQLLEIQPFARSFEMKEACYSATAALDY 155  
 IL+DED+ IDMVIVATSS IDQSKA++YVH L+ IQPFARSFEMK+K+ACYSATAALDY  
 35 Sbjct: 74 QILITDEKRAKIDMVIVATSSIDQSKAASVYVHQLLEIQPFARSFEMKEACYSATAALDY 133

Query: 156 AKLHVKEKHPDSKVLVIAASDIKYGKSTGSGTQAGSIAMLSQNFSLILELKEHDHLAQT 215  
 AKLHV PDS+VIAVIAASDI+YG+ S GESTQG+GSIA+L++ NP IL L ED++AQTR  
 40 Sbjct: 134 AKLHVASKPDSRVLVIAASDIKYGSGPGESTQSGSIALLVITANPRILALNEIDVIAQT 193

Query: 216 DIMDFWRPNYSIDVPYVNGMPSFKQYLDMLEKTIWKVYQKRPNTSLSDYAAFCFHIPPPKLA 275  
 DIMDFWRPNYS PTV+G++STKQYLL+ L+TIV+ YQKR N LSD AA CFHIPPPKLA  
 45 Sbjct: 194 DIMDFWRPNYSFTIPVDGIYSTKQYLINCLTETWQYQKREKQLSLDLAAVCFHIPPPKLA 253

Query: 276 LKGFNKLIDNNLDQKKARLQNFHSITYSKKIGNCTYGSILYLGLLSLENSQNLKAGD 335  
 LKG N I+DN + + + +L E F+ SI+YSK+IGN YTGSLYLGLLSLENS+ L++GD  
 50 Sbjct: 256 LKGLNNIMONTVPPEHREKLIRAFQASISYSKQIGNITYSGLYLGLLSLENSKVLQSGD 313

Query: 336 QIAFFSYSGGAVAEIPTQQLVDGYQNKILQSDRMQDLNKRQKLTIVTRYEKLFFFEKTLIDRN 395  
 +I FFSYSGGAV+E ++QQLV QY L ++R L++R +++V++YE LF+E+ LD+N  
 55 Sbjct: 314 KIGFFSYSGGAVSEFYSGQLVAGYDKMLNTRQALLDQRTLSVSKYEDLFYEQVGLDNN 373

Query: 396 GNANFNTRYTGTFSLDSICEHQRIY 420  
 GNANF+ Y TG F+L +I EH+RIY  
 55 Sbjct: 374 GNANFDLYLTKGFALTAIKHERRIY 398

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2165-

**Example 1923**

A DNA sequence (GBSx2032) was identified in *S.galactiae* <SEQ ID 5959> which encodes the amino acid sequence <SEQ ID 5960>. This protein is predicted to be HMG-CoA reductase (mvaA). Analysis of this protein sequence reveals the following:

5      Possible site: 50  
      >>> Seems to have no N-terminal signal sequence  
          INTEGRAL      Likelihood = -1.49      Transmembrane      348 - 364 ( 348 - 364)  
          INTEGRAL      Likelihood = -1.33      Transmembrane      53 - 69 ( 53 - 69)

10      ----- Final Results -----  
                  bacterial membrane --- Certainty=0.1595 (Affirmative) < succ>  
                  bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
                  bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

15      The protein has homology with the following sequences in the GENPEPT database.  
      >GF:AA02454 GB:AF290098 HMG-CoA reductase [Streptococcus pneumoniae]  
      Identities = 266/421 (63%), Positives = 343/421 (81%), Gaps = 3/421 (0%)

20      Query: 3      KISWTFPSKKSPERIHYLEQDFLADSSLEIVNQDLLSLSLANQMAENVIGRIALPFS 62  
                  KISW GFSKKS +ER+ L+ Q L+      + +S+++A+Q+ENV+G +LP+S  
      Sbjct: 2      KISMWFSKKSVCERLELLKQALLSPERQASLEKDBQMSVTVDQLSENVTGFSPLYS 61

25      Query: 63      LVPDVLVNGKVQVYFVTEEPSVVAASFAAKIIKRGGFLTVHNRKMIGQVALYVQD 122  
                  LVP+VLVNG+ Y VFYVTEEPSVVAAS+A+KIIKR+GGF VH R+MIGQVALY V +  
      Sbjct: 62      LVPEVLVNGQSVTVFVTEEPSVVAASYSKIIKRAGGFTAQVHQRMIGQVALYQVAN 121

30      Query: 123      SQETKESILNQKQQLLETANAHPISIVKRGGA CDL/TET--- KEDFLIVLVAVDTKEAM 179  
                  + +E I ++K +LLE+AN A+PSIVKRGGA DL +E + DFL+VY+ VET+EAM  
      Sbjct: 122      PKLAQEKIASKABLLLELANQAYPSIVKRGGA DLHVQKGEPEFLVVTYHVDTPQAM 181

35      Query: 180      GANMNVNMEALSSPLEDISKGSLSMISLNYATESLVATCRVDLFLSRQKEEAIKLA 239  
                  GANM+NTM+EAL LE++S+G+SLM ILSNYAT+SLVTA+CR+ R+LSRQK++ ++A  
      Sbjct: 182      GANMLNTMLEALKPVLEKLSQGSLSMGLISNYATDSLVTASCRIFAFLYLSRQKQGREIA 241

40      Query: 240      QKMTNLSQLAQDPYRASTHNGIPNGIDAIVLATGNDWRAIEAGHTYAVKDCQYRGLS 299  
                  +K+ +ASQ AQ DPYRA+THNGIPNGIDAI+ATGNDWRAIEAGH +A +DG+Y+GLS  
      Sbjct: 242      EKIALASQFAQADPYRAATHNGIPNGIDAILATGNDWRAIEAGAHAFASRDGRYQGLS 301

45      Query: 300      RWSYKVDNCLBGTLLFMPVATKGGSGINPSVHLAHDLLGRPNKELASIIISIGLAQ 359  
                  W+ ++ L G +TLEFMPVATKGGSGINP V L+HDLG P+A+ELA II+SIGLAQ  
      Sbjct: 302      CWTLLDERELVGBMTLMPVATKGGSGINLNRVALSHDLLGNPSARELAQIIVSIGLAQ 361

Query: 360      NFAALKALVSTGIQAGHMKLQAKSLALLAGAKBEQISEVVQQLDSKHMNLETAQKIVNKL 420  
                  NFAALKALVSTGIQ GHMKLQAKSLALLAGA E +++ +V++L+ K NLETAQ+ + L  
      Sbjct: 362      NFAALKALVSTGIQGHMKLQAKSLALLAGASESEVAPLVERLISDCTFNLETAQKIVNKL 422

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5961> which encodes the amino acid sequence <SEQ ID 5962>. Analysis of this protein sequence reveals the following:

50      Possible site: 31  
      >>> Seems to have no N-terminal signal sequence

55      ----- Final Results -----  
                  bacterial cytoplasm --- Certainty=0.3929 (Affirmative) < succ>  
                  bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                  bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 257/422 (60%), Positives = 330/422 (77%)

60      Query: 2      TKISWTFPSKKSPERIHYLEQDFLADSSLEIVNQDLLSLSLANQMAENVIGRIALPF 61  
                  T ++W+GFSKK+ EER+ +E+ L +L+ + LL+ ANQM ENV+GR+ALPF

-2166-

Sbjct: 4 TNLNWSGFSKKTFEERQLIKKFKLLNAENLNQLKTIVLLPIQTANQMTEENVIGRLALPF 63

Query: 62 SLVPPDLVNVGRYQVYFVVTTEPSVVAASFAAKIKRSGGLTTVHNRNIGQVALVDVQ 121  
S+ PD LVNG YQ+P+VTEEPSVVAASFAAK+IKRSGGF NR+NIQ+Q LND+

5 Sbjct: 64 SIAPDLVNGSTGYQMPFVTEEPSVVAASFAAKIKRSGGFKAQTINRIGQVIVNDID 123

Query: 122 DSQHTKESIANKQQLLEIANAAHPSIVKGGGACDIETIEKDFLIVYIMVDTKAMGA 181  
+ K +IL++ ++L+ +AN A+PSIVKGGGA + +E K +FLI YL VDT+KAMGA

10 Sbjct: 124 QIDNAGAILHRTKKLIALANKAYPSIVKGGGARTTHLEEGREFLIPYLTVDTKAMGA 183

Query: 182 NMVNIMMEALSSPLEDISKGSLSILSNYATESLVTATCRVDLRFLSRQKEATKLAQK 241  
NMVNIMMEAL L+SKG IM+ILSNYATESLVT +C +R L K +++LAQK

Sbjct: 184 NMVNIMMEALVPLDLRLSKGHCLMAILSNYATESLVTTSCEIPVRLIDHDKTKSLQLAQK 243

15 Query: 242 MTHASQLAQVDPYRATTHNGKIFNGIDAIVLATGNDWRATEGAHTYAVKDGQYRLSRK 301  
+ +AS+LAQVDPYRA+THNGKIFNGIDA+V+ATGNDWRATEGAH YA +G Y+GLS+W

Sbjct: 244 IELASRLAQVDPYRATTHNGKIFNGIDAVVIAIGNDWRATEGAHAYASRNGSYQLGLSQW 303

Query: 302 SYKVDNCLBGTLLTLPMPVATKGSIGINPSVHLAHDLLGRPNAKELASIIISIGLQNF 361  
+ D L G +TLMP+AR+KGSIG+NP+V +AHDLL +P+AK LA +I S+GLQNF

20 Sbjct: 304 HFDQKQVLGLQNTLRLMPIASKGSIGLNPVTSIAEDLLNQDAKTLAQIYASVGLQNF 363

Query: 362 AALKALVSTGIQAGHMKLQAKSLALLAGAKBEQISEVVKQLLDSKHMNLETAQKIVNKLT 421  
AALKAL S+GIQAGHMKL AKSLALLAGA +++I+ +V LL K +NLE A +++L

25 Sbjct: 364 AALKALTSSGIQAGHMKLAKSLALLAGATQDEIAPLVNALLADKPINLEKARFYLSQLR 423

Query: 422 KS 423  
+S

30 Sbjct: 424 QS 425

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1924

A DNA sequence (GBSx2033) was identified in *S. agalactiae* <SEQ ID 5963> which encodes the amino acid sequence <SEQ ID 5964>. Analysis of this protein sequence reveals the following:

Possible site: 52  
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.2355 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

45 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 5965> which encodes the amino acid sequence <SEQ ID 5966>. Analysis of this protein sequence reveals the following:

Possible site: 53  
>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.2687 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

55 An alignment of the GAS and GBS proteins is shown below.

Identities = 76/138 (55%), Positives = 100/138 (72%), Gaps = 2/138 (1%)

Query: 7 PKWEELPELLYLDQVLYVNLINPKITITNDKLTASMINNVYKHYIKSPIKKYKRR 66  
P W++LP+LDLYLDQVLYVNLQ + +++K LTASMINNVYKH Y++KPIKKY ++



-2167-

Sbjct: 7 PWKDLDPDLNLDQVLLVYNQCTDFSEVSDNKLSTASMINNYVGHVYTKPIKKYQKQ 66

Query: 67 QVARLIVITAPKQVFAIQBISQTELLTAINHSEAYNGFAACMNKEE--VHDLPPVVIS 124  
Q+ARLI I+ FK VF IQ+IS+ LE L A SR YN F C N++ D+PP+V

5 Sbjct: 67 QVARLIALSLFKVFPIQDISRVLLELQAQADSLSLYNTVTCWQKAPIEEDIPPIQV 126

Query: 125 ACOTILNYQETQKVLLEL 142  
ACQT+ Y +T L+ E+

10 Sbjct: 127 ACQTVKDYKTIYLLQEV 144

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1925

A DNA sequence (GBSx2034) was identified in *S. agalactiae* <SEQ ID 5967> which encodes the amino acid sequence <SEQ ID 5968>. This protein is predicted to be hemolysin iii. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

|    |          |                    |               |                        |
|----|----------|--------------------|---------------|------------------------|
| 20 | INTEGRAL | Likelihood = -9.08 | Transmembrane | 142 - 155 ( 140 - 165) |
|    | INTEGRAL | Likelihood = -6.79 | Transmembrane | 26 - 42 ( 19 - 44)     |
|    | INTEGRAL | Likelihood = -5.63 | Transmembrane | 200 - 216 ( 196 - 217) |
|    | INTEGRAL | Likelihood = -5.41 | Transmembrane | 104 - 120 ( 102 - 121) |
|    | INTEGRAL | Likelihood = -3.98 | Transmembrane | 51 - 67 ( 49 - 69)     |
| 25 | INTEGRAL | Likelihood = -1.86 | Transmembrane | 172 - 188 ( 169 - 188) |

----- Final Results -----

|    |                     |                                            |
|----|---------------------|--------------------------------------------|
| 30 | bacterial membrane  | --- Certainty=0.4630 (Affirmative) < succ> |
|    | bacterial outside   | --- Certainty=0.0000 (Not Clear) < succ>   |
|    | bacterial cytoplasm | --- Certainty=0.0000 (Not Clear) < succ>   |

A related GBS nucleic acid sequence <SEQ ID 9951> which encodes amino acid sequence <SEQ ID 9952> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:CAA58877 GB:X84058 novel hemolytic factor [Bacillus cereus]  
Identities = 79/204 (38%), Positives = 132/204 (63%), Gaps = 4/204 (1%)

Query: 17 EELANSITHAUGALLMLILLPITAVYSHRHFLQALGTSIFVTSFLPFLSSSIYHST 76  
EE+AN+ITH +G+L + L I +++ H A + +++ S+FL+L S++ H+

40 Sbjct: 14 EELANATHIGAILSLPALILILHASKHGTASAVVAFFVYGVSMFLYLFSTLLESH 73

Query: 77 YNSLQKYLVRMIDRSMIYIALAGSVTPVALSIGMLGLYLIIPLQWGITLGLILYKIFAP 136  
++K + ++DS IY+ IAG+YTP L + Q LG+ ++ W + + GI+KIF

45 Sbjct: 74 HKYVK-LPFDLSAISYLLGAGTYTTPFIIITRGLPWTLALINWFLATGSIIPKIFV 132

Query: 137 KINDKFSVLVHLNHWLVIP-IPFALITKGPAPWKLHLAGGICVTIGALFYA-RKRPVD 194  
+ K S + Y+IHWL+I I P TG F ILLAGGI Y++G+P+ K P++

50 Sbjct: 133 RRPFIKASTLCYITNGWLVITAKPLYNLTHGKF-SILLAGGILYSGVAIFPWLKILPFW 191

Query: 195 HMIHWLFPILLASHLQYIGIVYFVL 218  
H IHWLFL+L S + + ++++L

55 Sbjct: 192 HMIHWLFLVGGSAWFFCVLPVVL 215

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 5969> which encodes the amino acid sequence <SEQ ID 5970>. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

|    |          |                     |               |                        |
|----|----------|---------------------|---------------|------------------------|
| 55 | INTEGRAL | Likelihood = -10.51 | Transmembrane | 144 - 160 ( 138 - 163) |
|    | INTEGRAL | Likelihood = -9.87  | Transmembrane | 49 - 65 ( 45 - 71)     |

-2168-

INTEGRAL Likelihood = -7.11 Transmembrane 198 - 214 ( 193 - 215)  
 INTEGRAL Likelihood = -6.16 Transmembrane 102 - 118 ( 100 - 120)  
 INTEGRAL Likelihood = -2.97 Transmembrane 20 - 36 ( 20 - 41)  
 INTEGRAL Likelihood = -1.01 Transmembrane 167 - 183 ( 167 - 185)

5 ----- Final Results -----  
 bacterial membrane --- Certainty=0.5203(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the databases:

>GP:CAA58877 GB:X84058 novel hemolytic factor [Bacillus cereus]  
 Identities = 82/204 (40%), Positives = 128/204 (62%), Gaps = 4/204 (1%)

15 Query: 15 EEVANSVTHAIGAPAMLILLPIASAYAYQYDLKRAIGISIFVISLFIPLFSLSTIYHMA 74  
 EE+AN+TH IGA + L I +A + A + +++ +S+FL++L ST+ HS+  
 Sbjct: 14 EEIANVTHIGIGAILSPALILIIHASKGTASAVVAFTVYGVSMFLAYLFTLLHSTH 73

20 Query: 75 YGSVHRVILRIIDHSMIYIAGSYTPVALSLVSGMLGYIITVILQWGITLPGILYKIFAK 134  
 + V K + I+DHS IY+ IAG+YTP L + G LG+ ++ + W + + GI++KIF  
 Sbjct: 74 HPKVEK-LFTILDHSAIYLLIAGSYTPFLILITLGGPLGWTLLAIITWLAIGGIIFKIFV 132

25 Query: 135 RINEKPSIMLYVWGL-VVFIPLVITQKTSIAPGLMLPGGLSYTIGAVPIA-KRGPYP 192  
 R K S + YI+MGWL +V I P+ T F LL L GG+ Y++GA+P+ +K P+  
 Sbjct: 133 RRFIASTLCYIIMGWLIIAIVKPLYLEMTGHGFSLL-LAGGLISVGAIFFLWKLPPN 191

30 Query: 193 HMIWHLFLLASALQFIATFFML 216  
 H IWLHP+L SA+ F + F++L  
 Sbjct: 192 HAIWHLFVLGSSAMFFCFLPYVL 215

An alignment of the GAS and GBS proteins is shown below.

Identities = 153/213 (71%), Positives = 161/213 (84%)

35 Query: 6 SIKLSTQLSPGEEELANSITHVAGALLMLILLPIAVYSRHHFGLQALQTSIFVTSFLPM 65  
 + K S LSF EE+ANS+TH+GA MULLPIA Y++ + L+AA+G SIFV SLPLM  
 Sbjct: 4 TFRQSLPLSPGEEVANSVTHAIGAPAMLILLPIASAYAYQYDLKRAIGISIFVISLFIPLM 63

40 Query: 66 FLSSSIYHSMYNSLQYVLRMIDHSMIYIAGSYTPVALSLIGWGLYLIIFLQWGIT 125  
 FLSS+IYHSM Y S+ KY+K+IDHSMIYIAGSYTPVALSL+ GMLGY+II LQWGIT  
 Sbjct: 64 FLSTIYHSMAYGSEVHNILRIIDHSMIYIAGSYTPVALSLVSGMLGYIITVILQWGIT 123

45 Query: 126 LFGILYKIFKAPKINKPSLVLYLIMGWLVIIFAIITKIGPAFWGLLAGGICYTIGAL 185  
 LFGILYKIFA +IN+KPSI+LY++MGWL+PI P II KT AF L+L GG+ YTIAG+  
 Sbjct: 124 LFGILYKIFAKRINKPSIMLYVWGLVVFILPVITQKTSIAPGLMLPGGLSYTIGAV 183

Query: 186 FYAKRKPYPHMIWHLFLLASALQFIATFFML 218  
 FYA+KRPY HMIWHLFLLAS LQ+I I +FML  
 Sbjct: 184 FYAKRKPYPHMIWHLFLLASALQFIATFFML 216

50 Based on this analysis, it was predicted that these proteins could be useful antigens for vaccines or diagnostics.

### Example 1926

A DNA sequence (GBS:2035) was identified in *S.agalactiae* <SEQ ID 5971> which encodes the amino acid sequence <SEQ ID 5972>. Analysis of this protein sequence reveals the following:

55 Possible site: 52  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3641(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

SGP: CAB12492 GB: Z99107 similar to hypothetical proteins [Bacillus subtilis]  
Identities = 81/302 (26%), Positives = 157/302 (51%), Gaps = 10/302 (3%)

Query: 1 MKSAYIFPNPKGKDEQALAKEVKSYLEHDPQDDY-VRIITPSSVEEAVALKAKASEDH 59  
MK A I +NP SG++ + K+ + +++ Q Y + +A AK+A+  
Sbjct: 1 MKEARIIYNTPSGR---IPKHHLQVLQKPFQAGYETSTHATTCAGDATHAAKKAALRE 57

Query: 60 IDLVIPLGSDGTINKICGGVYAGGAYPTIGLVPAQTNNPNSKALNIPQERNL-ALENLNL 118  
DL+I GSDGTIN++ G+ PT+G++P GT N+P++AL IP+E L A + ++N  
Sbjct: 58 FDLIIAAGSDGTINEVNLGLADNRPTLGVIPVGTINDFARALGIPREDILQAADPTVIN 117

Query: 119 GHVKSVDICKVNDYMISSLTGLLADIAANVTSEMRRKLGPFAPLGDAYRLKKNRSYS 178  
G + +DI +VN Y I+ G L ++ +V S++K LG A+ +D R  
Sbjct: 118 GVAREFDIGQVNGQYFINIAGCGELTETYDVPSKLTMLGQLAYLKGMMELSLRPTE 177

Query: 179 IYLAYDNVRSRLRLLLLITMNSIAGMPAFSPPEATIDGFLRVYIMEHIHFLLHLR 238  
+ + YD + L L+T+TNS+ G +P++++DG+F + + + +  
Sbjct: 178 VSIEYDGLPQGEINLFLVTLTNSVGGTEKLAAPDSLNDGMFLMLIKCANLAEPIRVAT 237

Query: 239 QFRKGDFPSQAKEIKHFHTNNLTISTFKRKKSAPKVRIDGDPGDQLPVKVEVIPKALKFI 298  
+G+ + I + N + ++ ++ ++ +DG+ G LP + + + +  
Sbjct: 238 MALRGSHINDQHIIYTKANRVKVNVSFHM-----QLNLDGEYGNLPGFVNLKRIHV 292

Query: 299 IP 300  
+P  
Sbjct: 293 MP 294

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 5119> which encodes the amino acid sequence <SEQ ID 5120>. Analysis of this protein sequence reveals the following:

Possible site: 58  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4258 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 172/300 (57%), Positives = 229/300 (76%)

Query: 1 MKSAYIFPNPKGKDEQALAKEVKSYLEHDPQDDYVRIITPSSVEEAVALKAKASEDH 60  
MK+ IP+NP SGK E LA++VK Y +H F +D V++ITP ++A LAK+A++D I  
Sbjct: 1 MKTVRIFPNPNKSGKESQLARQVKDYFQCGPSEDSVKVITPDKADQAGFLAKQAKDI 60

Query: 61 DLVIPLGSDGTINKICGGVYAGGAYPTIGLVPAQTNNPNSKALNIPQERNLALENLNGH 120  
DLVIPLGSDGT+NKI GG+Y GGA+ IGLVP+GTVNNP+KA++IP + AL+ +L G  
Sbjct: 61 DLVIPLGSDGTINKIIGGYEGGAHCLIGLVPSTVNNPAKAMHPLPILQALDTILYTGQ 120

Query: 121 VKSV DICKVNDYMISSLTGLLADIAANVTSEMRRKLGPFAPLGDAYRLKKNRSYSIT 180  
+K VDICK N YMISSLTGLLADIAA+VT+E KR+ GP AFL D+ RLKKNRSY+I+  
Sbjct: 121 IKQVDICKANQYMISSLTGLLADIAADVTAERKRRGFLAPLAKDSIRILKKNRSYIS 180

Query: 181 LAYDNVRSRLRLLLLITMNSIAGMPAFSPPEATIDGFLRVYIMEHIHFLLHLRQF 240  
L N+ L+T+ LLITMNN+IAG P+PSP A DDG F+VYTM+ + FFK L H+ F  
Sbjct: 181 LISHNRHILKTKPLITMTNTIAGPSPSPSGAQLDGYFQVYTMKRVSPFKPLHINDF 240

Query: 241 RKGDFPSQAKEIKHFHTNNLTISTFKRKKSAPKVRIDGDPGDQLPVKVEVIPKALKFIIP 300  
++GDFSA+AEI HF N L++ +K++ +P+ RLXGD D LP++++IPKA+ I+P  
Sbjct: 241 KQGDPSKAEIISHFOANTLSLLPQAKQAILPRTRIDGSKSDYLPQLDIIPKAVSIIP 300

-2170-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1927**

A DNA sequence (GBSx2036) was identified in *S.agalactiae* <SEQ ID 5973> which encodes the amino acid sequence <SEQ ID 5974>. Analysis of this protein sequence reveals the following:

Possible site: 35  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3528 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB10885 GB:AB010693 gene\_id:K21C13.21-pir||T04769-strong  
similarity to unknown protein [Arabidopsis thaliana]  
Identities = 85/291 (29%), Positives = 150/291 (51%), Gaps = 28/291 (9%)

Query: 10 DQEWEPVBSGRYHMYGEPCFYAQRQIARQLGLD/KHISISFVDDV----- 57  
D + + P BGRYH+ + CP+A R ++ GLD+ I+ E V +  
Sbjct: 29 DPDSQFPBSGRYHLYISYACWACCLSYLKIKGLDSEAITFSSVHAIWGRKTETDDHRG 88

Query: 58 ----PSDIGLIFSQFEQVTKAKSLRDIVHLDTPTTYQGYPTIPIILLKNTNRIVCKESADI 113  
SD L ++P+ + GAKS-R++Y + P Y+G YT+P+L DK +V ES+++  
Sbjct: 89 WVFPSDTLPGASEFDYLNKAKSVRLYELASPNTEGKITYPVLWDKRLKKTWNESSEI 148

Query: 114 LRLFTTDPDSLHQNDAPVLFSQETASLINDIKDKINPQSLAYKLAFLDKQADYDTYSK 173  
+R+F T+P+ + + L+ +I+ + +YK P KQ Y+  
Sbjct: 149 IRMPNTEFNGIAKTPLSLDYPSLHRDVINETNGWVFGNGINGVYKOGFARKQEPYNEAVN 208

Query: 174 EFFTFLDQKEHLQRPPLLCNLSVDIHFTPLVRNDIAGRDILLNKKALEADYDYNIF 233  
+ + +D+ E +LG+ ++ G+ +E DI F L+R+D N++ L+YENIF  
Sbjct: 209 QLYEAVDRCEEVLGHRQRYTQNTFTADIRLFTLIRFDEVYAVIIFKCNKLLREYDYNIF 268

Query: 234 SWAKTLNDPMLKTI/TNPSIKNNFY----LGKPGRAVRHHTIVPTGPNM 279  
++ K +Y + + N + IK +YY + PG I+P GPN+  
Sbjct: 269 NYIKDIYQLHGMSSVTNMEHIKQHYGSHPTINPFG-----IIPGPNL 312

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1928**

A DNA sequence (GBSx2037) was identified in *S.agalactiae* <SEQ ID 5975> which encodes the amino acid sequence <SEQ ID 5976>. Analysis of this protein sequence reveals the following:

Possible site: 59  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2647 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB07793 GB:AB037666 hypothetical protein [Streptomyces sp.  
C11901]

-2171-

Identities = 127/331 (38%), Positives = 194/331 (58%), Gaps = 9/331 (2%)

Query: 4 RKDDHIKALKYQSHY--NSFDDIELIHSSLPKYNVDIDLSTHFAGQSFRFFPYINAM 60  
 5 Sbjct: 6 RKDDH+LAIEQHNAHSGRNQFDVSVFHHLAGIDRPDVSLSATSFAGISWQVPIYINAM 65

Query: 61 TGGSEKGAHVNHKLAQVAQATGIVMVTGYSAAALKNDE--DSYPTTDLYPDILKATINIG 118  
 10 Sbjct: 66 TGGSEK +N LAA A+ TG+ +GS +A +K+ D D P+ + NI TGGSEKTLNRDLATAARETGVPIAGSMNAYIKDPSCAITFRVLKIDENPNFVIANIN 125

Query: 119 LDKPVPAARSTVKAMNPFLQVHVNVQMQLMPEGREREFHMRSHLAKYVDINICPLILK 178  
 20 Sbjct: 126 ATTTVDNAQRAIDLEIANALQIHINTAQETPMPEGRSFAWVQPIKIAAAVDIPVIVK 185

Query: 179 EVGFGMDLQSIKDAYDIGITTTVDISGRGTSFAYIENQRGR--DRSYLNTWQQTAAQSLI 236  
 25 Sbjct: 186 EVNGLSRCTLILLADLGVAQADVSGRGCTDFARTENGRRELGDYAFHGWGQSTAAACLL 245

Query: 237 NAQSMQDKMDILASGGIRHPLDMVKCLVLAQAVGLSRTVLELVERYPVDDVIAILNSWK 296  
 30 Sbjct: 246 DAQDII--SLPVLASGGVRFHLDVVRALALGARAVGSSAGFLRLTMDGDVALITKLTITWL 303

Query: 297 EDLRMINCALNCKKITDLRQVNYILYQQLKE 327  
 35 Sbjct: 304 DQLAALQTMLGARTPADLTRCDVLLHKELRD 334

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5977> which encodes the amino acid sequence <SEQ ID 5978>. Analysis of this protein sequence reveals the following:

Possible site: 51  
 30 >> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2823 (Affirmative) < succ>  
 35 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 244/329 (74%), Positives = 284/329 (86%)

40 Query: 1 MTRKDDHIKALKYQSHYNSFDDIELIHSSLPKYNVDIDLSTHFAGQSFRFFPYINAM 60  
 Sbjct: 31 MTRKDDHIKALKYQS YN+FDIELIH SLP Y++DIDLSTHFAGQ F+FFPYINAM 90

45 Query: 61 TGGSEKGAHVNHKLAQVAQATGIVMVTGYSAAALKNDEDSYPTTDLYPDILKATINIGLD 120  
 Sbjct: 91 TGGSGKGAHVNHKLA+VA ATGIVMVTGYSAAALKN DSY ++ +LKATINIGLD 150

50 Query: 121 KPVPAARSTVKAMNPFLQVHVNVQMQLMPEGREREFHMRSHLAKYVDINICPLILKEV 180  
 Sbjct: 151 KPVALQQTVCQMOPFLQVHVNVQMQLMPEGRERVFHTWKHLARYASQIPVPVILKEV 210

55 Query: 181 GFGMDLQSIKDAYDIGITTTVDISGRGTSFAYIENQRGRDRSYLNTWQQTAAQSLINAGS 240  
 Sbjct: 211 GFGMDVNSIKLAHDIGITTTVDISGRGTSFAYIRNQRGGDRSYLNTWQQTTCVCLINAGS 270

60 Query: 241 MMDKMDILASGGIRHPLDMVKCLVLAQAVGLSRTVLELVERYPVDDVIAILNSWKDLR 300  
 Sbjct: 271 LMDQVBIASGGVRFHPLDMKICFVLGARAVGLSRTVLELVEKYPTERTVIAIVNGWKEKLR 330

60 Query: 301 MIMCALNCKKITDLRQVNYILYQQLKAN 329  
 Sbjct: 331 IIMCALDCKITIKELKGVVDYILYQLQVN 359

-2172-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1929

A DNA sequence (GBSx2038) was identified in *S. agalactiae* <SEQ ID 5979> which encodes the amino acid sequence <SEQ ID 5980>. This protein is predicted to be phosphomevalonate kinase. Analysis of this protein sequence reveals the following:

Possible site: 41  
>>> Seems to have no N-terminal signal sequence

#### ----- Final Results -----

bacterial cytoplasm --- Certainty=0.0785 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AG02457 GB:AF290099 phosphomevalonate kinase [Streptococcus pneumoniae]  
Identities = 170/330 (51%), Positives = 233/330 (70%), Gaps = 1/330 (0%)

Query: 1 MVKVGKGLYIAGEYAILYPGQVAILKNVPIYMTALATFADNYSLSYDMFHTASLQPD 60  
M+ V+T GKLY AGEYAIL PGQ+A++K++PIYM A F+D+Y +YSDFM++ L+P+  
Sbjct: 1 MIAVTCGKLYNAGEYAILRPGQALIKDIPYMRABIAFSDSYRISYDMFDFAVDLRPN 60

Query: 61 KQYSLIQETILLMEENLINFQGNIKPIHLEITGKLERYGLKFGIGSGSVVVLTIKAA 120  
YSLIQETI LM ++L G+N++P L+I GK+ER G KFG+GSSGSGVVVL +K+ A  
Sbjct: 61 PYSLIQETIALMGDFLAVRGQNLRFPSLKTCHNREGRKFKLGSSGSGVVVLVVKALL 120

Query: 121 LYEIRMFSDLLFKLSAYVLLKRGDNGSMGDIACTAYEHLISYAFDRFRAVSKMIETKPLE 180  
LY ++ +LLFKL++ VLLKRGDNGSMGD+ACT E L+ Y +FDR+ + +E + L  
Sbjct: 121 LYNLSDVQNLFLTSBAVLLKRGDNGSMGDIACTAVRDVLVQSGDRQKAAMLESEKLA 180

Query: 181 QVLEAEKGYRITKIQALLEMDPLVGWMTQPSISKEMINIVKSTITRFLDDTKQVQVLL 240  
VLE +WG+ I++++ LE DFLVGMT + ++S M+ +K I Q FL +K VV L+  
Sbjct: 181 TVLERDWGFPISQVKPTECDFLVGMTKEVAVSSHVVQIQKINQNLSSSKETVSVL 240

Query: 241 SAKFKDQKEAIKRCLEEISLLPFLHPSITYTDKLQKESKGLDIVTKSSGSGGGDCGI 300  
A ++G E + +E S LL L IYT L++LKEAS+ L V KSSG+GGGDCGI  
Sbjct: 241 EALEQQAQAEKVTEQVEVASKLEGLSDTYTFLRLQKESQDLQAVAKSSGAGGGDCGI 300

Query: 301 AISPFI-KNDNQTLIKRWESAGIELLSKETL 329  
A+SF+ ++ TL RW GIELL +E +  
Sbjct: 301 ALSFDQSSRNVLKRWADLQIELLYQERI 330

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 5981> which encodes the amino acid sequence <SEQ ID 5982>. Analysis of this protein sequence reveals the following:

Possible site: 59  
>>> Seems to have no N-terminal signal sequence

#### ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2669 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 171/325 (52%), Positives = 227/325 (69%), Gaps = 2/325 (0%)

Query: 4 VQTGGKLYIAGEYAILYPGQVAILKNVPIYMTALATFADNYSLSYDMFHTASLQPDQKY 63  
VQTGGKLY+ GYEAAIL PGQ A++ +P+ MTA + A + L SDMP++ A + FD Y  
Sbjct: 22 VQTGGKLYIAGEYAILTPGQALIHFIPIAMTARISPAHIQIASDMPFHKAGMTPDASY 81

-2173-

Query: 64 SLIQETILMEEWLINFGRNKKPIHLBITGKLBRYGLKFGIGSSGSVVVLTIKMAALYE 123  
 +LIQ T+ ++L' ++P L ITGK+ER G KFGIGSSGSV +LT+KA+A Y+  
 Sbjct: 82 ALIQATVTKTFADYLGQSIDQLPEPFLITGKMERDGGKFGIGSSGSVTLTLKALSAYYQ 141

Query: 124 TEMPSDLLFKLSAYVLLKRGDNGSMGDIACTAYEHLISYSAFDRRAVSOMIETKPLEQVL 183  
 I + +LLFKL+AY LLK+GDNGSMGDIACTAY+ L++Y++FDR VS ++T PL++LL  
 Sbjct: 142 ITLTPELLFKLAAYTLKQGDNGSMGDIACTAYQTLVATTSFDRBQVSNLQMPLEKLL 201

Query: 184 EAEWGYRITKIQALLMDPLVGMWQPSISKEMINIVKSTTIFRFLDDTKYQVQVQ-LLSA 242  
 +WGY I IQ L DFLVGMT P+IS++MI V ++IT FL T YQ+ Q +A  
 Sbjct: 202 VKDWGYHIQVIQPALPCDFLVGMTKIDPAISRQNIQQVTASITPAFL-RTSYQLYQSAWVA 260

Query: 243 FKEGDKEATKRCLEISILLNLEPSIYTDKIQKLRASKGLDITVKSQSGGGGDCGIAI 302  
 +EG KS +K+ L S LL LHP+IY KL L A + D V KSSQSGGGGDCGIAI+  
 Sbjct: 261 LQBGHKEELKKSLAGASHLLKLELPAITYHPKLVTLVAACQKQDAVAKSSQSGGGGDCGIAL 320

Query: 303 SFHQNINQYTLIKRMESAGIELLSKE 327  
 +FHN++ TLI +W+ A I LL +E  
 Sbjct: 321 AFHQDARDTLISKWQADIALLYQE 345

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1930

- 25 A DNA sequence (GBSx2039) was identified in *S. agalactiae* <SEQ ID 5983> which encodes the amino acid sequence <SEQ ID 5984>. Analysis of this protein sequence reveals the following:

Possible site: 32  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -1.75 Transmembrane 20 - 36 ( 18 - 36)

30 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1702 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

35

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### 40 Example 1931

A DNA sequence (GBSx2040) was identified in *S. agalactiae* <SEQ ID 5985> which encodes the amino acid sequence <SEQ ID 5986>. This protein is predicted to be mevalonate diphosphate decarboxylase. Analysis of this protein sequence reveals the following:

Possible site: 25  
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1557 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

50

The protein has homology with the following sequences in the GENPEPT database.

>GP:ANG02456 GB:AF290099 mevalonate diphosphate decarboxylase  
 [Streptococcus pneumoniae]

-2174-

Identities = 219/312 (70%); Positives = 264/312 (84%)

Query: 1 MDGKSISVKS YANIAIIKYWGKADAKRMIPATSSISILLENMYTETRI/TALGKDAKDEF 60  
 MD ++V+SYANIAIIKYWGK ++M+PATSSISILLENMYTET L+ L + DEF  
 5 Sbjct: 1 MDREPTVTSYANIAIIKYWGKKKKRNVPATSSISILLENMYTETLSP/LANVTADEF 60

Query: 61 YISGVLQNDHEHDIMSAILDPRQNRSGFVKIETNNMPTAAGLSSSSSGLSALVACND 120  
 YI+G LQN+ EH KMS I+DR+R GPV+I+T NNMPTAAGLSSSSSGLSALVACN  
 10 Sbjct: 61 YINQQLQNEVEHAKMSKTIIRYRPAQRGPVRIDYNNMPTAAGLSSSSSGLSALVACNA 120

Query: 121 FPGTHLSQSQAQAKFASGSSSRSPFGPVAAMWKKDSCDIYKVHTNLDLAMIMLVNDRK 180  
 +F L +SQAQAKFASGSSSRSP+GP+ AMDKDG+IY V T+L LAMIMLVN DR+  
 15 Sbjct: 121 YFKIGLDRSQAQAKFASGSSSRSPYGLPVAAMWKKDSCRIY PVETDLKIMLVNLEDKK 180

Query: 181 KPISSREGMKICTETSTTFNEMVRQSEQYQDMVLVYKNDPQKVQQLTERNALMHSITT 240  
 KPISSR+GMK+C ETSTTF++WVRQSE+DYQDM+LYK NDF K+G+ITE+NALAMH+TT  
 20 Sbjct: 181 KPISSRDGMKLCVETSTTFDQWVRQSEKQYQDMVLVYKNDPAKIGELTERNALMHSITT 240

Query: 241 KTATPAFSYLTETTYKAMDVVKKLREKGEHCYTTMDAGPNVKVLCRLQDLAALAEK 300  
 KTA+PAFSYLT+ +Y+AM V++LREK GY+TMDAGPNVKV C +DL L+ I +  
 25 Sbjct: 241 KTAPAFSYLTASYTEAMAFVRLQREKGEACYTTMDAGPNVKVFCOEKLEHLSEIFQGR 300

Query: 301 YRIIVSTTKELA 312  
 YR+IVS TK+L+  
 30 Sbjct: 301 YRLIVSKTKDLS 312

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5987> which encodes the amino acid sequence <SEQ ID 5988>. Analysis of this protein sequence reveals the following:

Possible site: 36  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1271 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 35 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 221/313 (70%), Positives = 258/313 (81%)

Query: 1 MDGKSISVKS YANIAIIKYWGKADAKRMIPATSSISILLENMYTETRI/TALGKDAKDEF 60  
 +D I+V SYANIAIIKYWGK + KMIP+TSSISILLENM+T F ++ L A D+F  
 Sbjct: 1 VDPNVITVTSYANIAIIKYWGKQAKMIPATSSISILLENMYTETTSVFLPDATSDQF 60

Query: 61 YISGVLQNDHEHDIMSAILDPRQNRSGFVKIETNNMPTAAGLSSSSSGLSALVACND 120  
 YI+G+LQND EH K+SAI+D+FRQ FVK+ET NNMPTAAGLSSSSSGLSALVAC+  
 45 Sbjct: 61 YINGILQNDHEHMKISAIIDQFRQPGQAFVKMETNNMPTAAGLSSSSSGLSALVACQ 120

Query: 121 FPGTHLSQSQAQAKFASGSSSRSPFGPVAAMWKKDSCDIYKVHTNLDLAMIMLVNDRK 180  
 F T L Q LQ+AKFASGSSSRSPFGPVAAMWKKDSCG IYKV T+L +AMIMLVN +  
 50 Sbjct: 121 LFDQQLDQKALQAKAFASGSSSRSPFGPVAAMWKKDSCGAIYKVETDLMAMIMLVNAAK 180

Query: 181 KPISSREGMKICTETSTTFNEMVRQSEQYQDMVLVYKNDPQKVQQLTERNALMHSITT 240  
 KPISSREGMK+C +TSTTF++WV QG DYQ ML YLK N+P+KVGQLTE NALAMH+TT  
 55 Sbjct: 181 KPISSREGMKLCRD+STTFDQWVRQSAIDYQMLTYLKTNNPEKVQQLTERNALMHSITT 240

Query: 241 KTATPAFSYLTETTYKAMDVVKKLREKGEHCYTTMDAGPNVKVLCRLQDLAALAEK 300  
 KTA P FSYLT+H+Y+AM+ VK+LR+G CY+TMDAGPNVKVLCRL+DL LA L K+  
 60 Sbjct: 241 KTANPPFSYLTESYQAMEAVKELRQEGFACYTTMDAGPNVKVLCRLQDLAALAEK 300

Query: 301 YRIIVSTTKELAD 313  
 YRIIVS TK+L D  
 Sbjct: 301 YRIIVSKTKDLPD 313



-2175-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1932

A DNA sequence (GBSx2041) was identified in *S. agalactiae* <SEQ ID 5989> which encodes the amino acid sequence <SEQ ID 5990>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1512 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 5991> which encodes the amino acid sequence <SEQ ID 5992>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1117 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 182/290 (62%), Positives = 223/290 (76%)

Query: 1 MKKKFGIGKAHSHKIIIMGEHVVYGYPAIAIPKNIIEVTCLEERAPQILALMDTDLSTA 60
M E G GKAHSHKIIIMGEHVVYGYPAIAIPKNIIEVTCLEERAPQILALMDTDLSTA 60
Sbjct: 6 MGENIGYGAHSHKIIIMGEHVVYGYPAIAIPKNIIEVTCLEERAPQILALMDTDLSTA 65

Query: 61 IFALDYLGKTSKIAIYHISCVFERGMGSSAAVAIAIAIRAVFOYFDELDLEADLEICLV 120
I+A+LDYL + IAY I SQVP++RGMGSSAAV+IAIAIRAVF Y E L DLE LV
Sbjct: 66 IYASLDYLRQLQEPYIEIVSQVFQKRGMGSSAAVSIATRAVFSYQCEPLSDLLRLIV 125

Query: 121 NRAEMIAHNSPGLDAKTKLSENTIKFIRNIGFSTVPHLMAYLVIAIDTGIHGHTKEAVD 180
N+AE+IAH+NPSGLDAKTKL++ IKFIRNIGF T+ + LN YL+IADTGIHGHT+EA+V
Sbjct: 126 NKABTIAHNPGLDAKTKLSDHAKFIRNIGFETIEIALNGYLIADTGIHGHTREAVN 185

Query: 181 KVKSQGEAVLPLKELGYLASEDAHKSDSKQLGSLMTKAHOSLKLGLVSSLEADHLV 240
KV E LP+L +LG L +A E AI++ + +G LMT+AH +LK +GVS +AD LV
Sbjct: 186 KVAQFEETNLFLYLAKLGLALQALERAINQKNAIGQLMTQAHSALKAIGVSIKADQLV 245

Query: 241 EVAISGALGAKMSGGGLGSCIALVKEKREARELSQQLEREGAVNTIWE 290
E A+ GALGAKM+GGGLGSC+IAL K AE++S +L+ EGAVENTW +
Sbjct: 246 EALRAGALGANVTGGGLGSCMIADTKDMAEKISIRKEEGAVNTIWQ 295
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1933

A DNA sequence (GBSx2042) was identified in *S. agalactiae* <SEQ ID 5993> which encodes the amino acid sequence <SEQ ID 5994>. This protein is predicted to be a histidine protein kinase. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -13.43 Transmembrane 12 - 28 (4 - 33)
```

-2176-

INTEGRAL Likelihood = -9.29 Transmembrane 163 - 179 ( 157 - 191)

----- Final Results -----

- 5 bacterial membrane --- Certainty=0.6371 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 10 >GP:AAF79919 GB:AF039082 putative histidine protein kinase  
 [Lactococcus lactis]  
 Identities = 78/315 (24%), Positives = 154/315 (48%), Gaps = 33/315 (10%)
- Query: 101 SDRQIKNYAKRIVSQNSHSGHITINFSTYSYLKKVGNKNDLVVFLDTINQYLDNRLLQ 160  
 +++QI N + + +N+ + Y + T S + V+ + + Q +  
 15 Sbjct: 84 NEEQI-NTIQT-VRNPNYGINWHYRLTTSQFIITNSDGTPTVPVQIFSNVDQIQDMS 142
- Query: 161 LSIWM---SLVSPFIVFMVIVSV-LSGRVILFPVANYEKQRRFITNAGHELKTPLAISAN 216  
 ++N+ +++P + VI+S+ L+ + P +A YEKQ+ P+ NA HEL+TFLA+  
 20 Sbjct: 143 RAMVIVVTMITTFWILSVIISLVIANWILKPIAAEYKQKEFFVNASHELRTFLAILQNR 202
- Query: 217 NELV-----EMMSGESEWTKSTNDQIQLTGILNGMVSLAR-----FERQPDISM--- 261  
 EL+ + +SE + +++ L + ++LAR E +P +  
 20 Sbjct: 203 LELLFQKPTATIIDQSENISESLSEVRNMRLTSLNLINLARDDSGIKIEPEPTATYFEN 262
- Query: 262 VDLDFSHITKDAEDFKGPIIKDGKDFINSIQPGIHVKAEBKSELPVLVLVDNANKYCD 321  
 + + +TAA+ + F G + +G V ++ + +L+T+L DNA KY D  
 25 Sbjct: 263 IFNSYEMLTENAGKKFSGNLKLEGT-----VNLDQALIKQLTILFDNALKYTD 311
- Query: 322 FMGTIVTVLGRSSRLRRKLEVSNVYQNGKIDIDSKPFFRYFEDESHNKKSGYIGLS 381  
 G ++V + ++ V++ + D D K F+R+R D++ +K G+GLS  
 30 Sbjct: 312 SEGEISVDVINNGGF--LTPAVADNGEGISDEDEKKKIFRFRFVDKATRKQKGLGLGLS 369
- Query: 382 IVTSLVHLFKGSDIV 396  
 + +V + G I V  
 35 Sbjct: 370 LAKQIVEAYNGKITV 384

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5751> which encodes the amino acid sequence <SEQ ID 5752>. Analysis of this protein sequence reveals the following:

- 40 Possible site: 24  
 >>> Seems to have an uncleavable N-term signal seq
- INTEGRAL Likelihood = -11.30 Transmembrane 18 - 34 ( 13 - 42)  
 INTEGRAL Likelihood = -10.35 Transmembrane 170 - 186 ( 163 - 199)
- 45 ----- Final Results -----
- bacterial membrane --- Certainty=0.5522 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

- 50 Identities = 233/410 (56%), Positives = 303/410 (73%), Gaps = 1/410 (0%)
- Query: 1 MFNRLRIRFGIAALAILVFLPSVGVGLNSANEYQTKNEIRVRLTILADRNGRIPNKLFP 60  
 MF +R+RFI IA++AI ++L S+VG++N+A YQ++ EI R+L +++ N G++P E  
 55 Sbjct: 10 MFNRLRIRFIMIASIAIFILSSVIGIINTARCYQSQEIINRLLELISNNKGLPOTTES 69
- Query: 61 SKELGDLSTDAIPQFRYFSARTAKGNVTSFDSRNI FEVSDROIKNYAKRIVSQNSHSG 120  
 SK LG LS D++ QFRY+S +A G++ S ++ NI + + + +A+  
 55 Sbjct: 70 SKRLCTKLSSELSQFRYYSVIPNAGHLSSNTNITSGALDRBAQYFARLFAKSGEKEG 129
- Query: 121 HITYNFSTYSYLKKVGNKNDLVVFLDTINQYLDNRLLQLSISWMSLVSPFIVFMVIVSVL 180  
 + S YSYL+ ++ + LUV LDTT + LL +S+ ++ FI P+V+VS+  
 60 Sbjct: 130 SYRHQDSVYSYLITQLPNEEKLVVILDTTFYFRSVGDLIAVSVMLAFGGFIFFVVLVSLF 189
- Query: 181 SGRVILFPVANYEKQRRFITNAGHELKTPLAISANNELVEMMSGESEWTKSTNDQIQL 240

-2177-

SG VI PFV NYRQRRFITNAGHELKTLAIISANNELVE+M+GSESWTKST+DQ++RL  
 Sbjct: 190 SGMVIKPFVQNYRQRRFITNAGHELKTLAIISANNELVELMTGSESWTKSTSDQVKRL 249

Query: 241 TGLINQVNLARLAEQPDISMVDLPFHSITKDAARDFKGPIIKDKGDFIMSIQPGIRVKA 300  
 5 TGLIN M++LAR EQPD+ + +DPS I +DAARDFK ++KDGK F ++IQP I +KA  
 Sbjct: 250 TGLINQMTTLARLAEQPDVVLHMVDFSAIAQDAARDFKSLVLKDGKRFDLTIQENIMIKA 309

Query: 301 ESKSLFELVTLVLNANKYCDPMGTIVTVKLSRSLR-RAKLEVSNTYKNGKDIDYSKFP 359  
 ESKSLFELVT+LVLNANKYCDP G V V L+ R R RAKLEVSNTY GK IDYS+FF  
 10 Sbjct: 310 ESKSLFELVTLVLNANKYCDPKGLVKVSLTTIGRRRKRAKLEVSNTYLGKSIDYSRFF 369

Query: 360 ERFYREDESHNNKSGYIGLSIVTSLVELPKGSIDVNYKHDITTFVIYI 409  
 ERFYREDESHN+K+ GYIGLS+ S+V LFKG+I VNYK+D I F + I  
 15 Sbjct: 370 ERFYREDESHNSKEGYIGLSMAKSMVKLPKGTITVNYKNDIVFTVVI 419

SEQ ID 5994 (GBS273) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 14; MW 46kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 56 (lane 5; MW 71kDa).

GBS273-GST was purified as shown in Figure 208, lane 4.

- 20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1934

A DNA sequence (GBSx2043) was identified in *S. agalactiae* <SEQ ID 5995> which encodes the amino acid sequence <SEQ ID 5996>. Analysis of this protein sequence reveals the following:

25 Possible site: 16  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2181 (Affirmative) < succ>  
 30 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1935

A DNA sequence (GBSx2044) was identified in *S. agalactiae* <SEQ ID 5997> which encodes the amino acid sequence <SEQ ID 5998>. This protein is predicted to be two-component response regulator (trcR).

- 40 Analysis of this protein sequence reveals the following:

Possible site: 56  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 45 bacterial cytoplasm --- Certainty=0.2503 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 50 A related GBS nucleic acid sequence <SEQ ID 9379> which encodes amino acid sequence <SEQ ID 9380> was also identified.

-2178-

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA004091 GB:AP001508 two-component response regulator [Bacillus halodurans]  
Identities = 71/183 (38%), Positives = 120/183 (64%), Gaps = 3/183 (1%)

5 Query: 9 RVLIAEDDEEQMSRVLSTAIHQGVVVDVAYDQQTADLANQAYDVMMVDMVPVKTGIE 68  
R+LI EDE+++RVL + H+GY D A+ G ++ +A+D+++DVM+P +G+E  
Sbjct: 3 RILIIIDEKKIARVLQLELSEHGSETDAAFSGSGLETTQAHAWDLVLLDVMLEPESGLE 62

10 Query: 69 AVKEIRQSGNKSIIIMLTAMAEIDRVVTGLDAGADDYLTKEPFLKELLARLRSMRRLE 127  
++ IR + + II+LTA I D+V+GLD GA+DY+TKPF ++ELLAR+R ++  
Sbjct: 63 VLRIRMTDPVTFPIILLTARNIPDKVSGLDIGANDYITKPFLEKELLARVRACLRTVQT 122

15 Query: 128 DFTFNVLISGRVTLVGQELQECN-TIRLAGKEAKMLAFFMLNHDKELSTQQLFEHVW 185  
+ + L +T++ ++Q N TI L KE ++L FF+ N + LS +Q+ +VW  
Sbjct: 123 RERVETLMFQELTINENKRDVQRGNSTIELTKPFELIAVFFIKNQQLVRSQILTNVW 182

Query: 186 GAC 188  
G C  
Sbjct: 183 GFD 185

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5999> which encodes the amino acid sequence <SEQ ID 6000>. Analysis of this protein sequence reveals the following:

Possible site: 21  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2391 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 125/185 (67%), Positives = 151/185 (81%)

35 Query: 8 MRVLIARDEEQMSRVLSTAIHQGVVVDVAYDQQTADLANQAYDVMMVDMVPVKTG 67  
N++L+AEDE QMS VL+TA++HQGY VDV ++GQ AID A NAYD+M++D+MMP+K+GI  
Sbjct: 1 MKILLARDEEQMSNVLTITANTHQGYDVDFVFGQSAIDKAKONAYDITMILDMPIKSGI 60

40 Query: 68 EAVKEIRQSGNKSIIIMLTAMAEIDRVVTGLDAGADDYLTKEPFLKELLARLRSMRRLE 127  
EA+KEIR SGN SHIIMLTAMAEI+DRVVTGLDAGADDYLTKEPFLKELLARLRSM RR+E  
Sbjct: 61 EALKEIRASGNCSIIIMLTAMAEINDRVVTGLDAGADDYLTKEPFLKELLARLRSMRRVE 120

Query: 128 DFTFNVLISGRVTLVGQELQECNTIRLAGKEAKMLAFFMLNHDKELSTQQLFEHVWGA 187  
FTT VT. VTT++ RORT. N TTA KE K++P MTA K T. TA T++G

-2179-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:BA05604 GB:AP001513 unknown conserved protein [Bacillus halodurans]  
 Identities = 67/182 (36%), Positives = 111/182 (60%), Gaps = 4/182 (2%)  
 Query: 17 LEDFSQRIQLNDKAKVETGYKLYEHIIGRIKTSDSMIKCRRKQLPVTVDALKTIRDS 76  
 L++ + +I + + + Y EH+ R+K+ +S++ K +R+ T++S + +RD  
 10 Sbjct: 29 LQELNTKIDILKQEFQYTHDYNPIEHVSSRVKSPESIVNKIQRRGNDFTLESIRNVRDI 88  
 Query: 77 IGVRLICGFVNDIYQIIRIKAFDDCRIVVEKDYIQHVKPNGYRSYHVILEIDTFYPDCL 136  
 G+RI C F +DIY + E++ D +V KDYI++ KFNYSR H+IL I P +  
 15 Sbjct: 89 AGIRITCSFSDIYTLSEQLMGCHDISVVETKDYIKNPKPNGYRSLHLILSI---PTFM 144  
 Query: 137 GNSDGKYYIEIQLRTIAQDSWASLEHQMCKYKHDIENPERIVRELKRCADENASVDLTMQT 196  
 + Y+E+Q+RTIA D WASLEH++ YK++ PE +++ELK A+ A +D M+  
 Sbjct: 145 SDRVQDVYVEVQIRTIAMDPFWASLEHKIYYKYNKNVPEHLKELKDAESAALDQKREK 204  
 20 Query: 197 IR 198  
 I+  
 Sbjct: 205 IQ 206

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6003> which encodes the amino acid  
 25 sequence <SEQ ID 6004>. Analysis of this protein sequence reveals the following:

Possible site: 33  
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1057 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

35 Identities = 127/206 (61%), Positives = 162/206 (77%)  
 Query: 3 TNIYGDYGRYLPILLEDFSQRIQLNDKAKVETGYKLYEHIIGRIKTSDSMIKCRRKQL 62  
 ++IY + YLPL+L+ + I EN K+K ETG+KLYEH RIK+ SMIEKC+RKQL  
 40 Sbjct: 11 SSIYSGFZYVLPLVLQTTIDVIAENIKSKSGTGPFLYEHFTSRISSEASMIKCCGRQL 70  
 Query: 63 PVTVDSALKTIRDSIGVRIICGFVNDIYQIIRIKAFDDCRIVVEKDYIQHVKPNGYRSY 122  
 P+T SALK I+DSIG+RIICGF++DIY++++ +K+ + EKDYI + KPNGYRSY  
 45 Sbjct: 71 PLTKSALKIIRDSIGIRIICOFIDDIYRMVDLLKSPGMSVTEKDYILNAKPNGYRSY 130  
 Query: 123 HVILEIDTFYPDCLGNSDGKYYIEIQLRTIAQDSWASLEHQMCKYKHDIENPERIVRELKR 182  
 H+ILE++T+D LG G Y+IE+QLRTIAQDSWASLEHQMCKYKH + N E I RELKR  
 50 Sbjct: 131 HLILELETHFPDILGKNGCYPIEVQLRTIAQDSWASLEHQMCKYKHQVANAMITRELKR 190  
 Query: 183 CADENASVDLTMQTIRQLIRSGTKKE 208  
 CADE+AS D+TMQITRQLI+ T++E  
 55 Sbjct: 191 CADELASCIVTMQITRQLIQETIKER 216

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 vaccines or diagnostics.

### 55 Example 1937

A DNA sequence (GBSx2046) was identified in *S.agalactiae* <SEQ ID 6005> which encodes the amino  
 acid sequence <SEQ ID 6006>. Analysis of this protein sequence reveals the following:

Possible site: 40  
 >>> Seems to have no N-terminal signal sequence

-2180-

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3250(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAI37193 GB:X53013 ORF1 (AA 1 - 384) [Lactococcus lactis]  
 Identities = 30/55 (54%), Positives = 37/55 (66%)

Query: 1 MSFYKTLRKRFINDADTIFIEQSQFEIPFIYIETDHNSSSHVLDYQSQKFEK 55  
 ME +YKTLRR+ INDA ++ EIF YIET +N+ H LDYQS K+FEK  
 Sbjct: 327 MSFYKTLRKRLINDAHFETRASATQEIFKYIETYNIGMHSGLDYSPKDFEK 381

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6007> which encodes the amino acid sequence <SEQ ID 6008>. Analysis of this protein sequence reveals the following:

Possible site: 36  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3055(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 31/59 (52%), Positives = 39/59 (65%)

Query: 1 MSFYKTLRKRFINDADTIFIEQSQFEIPFIYIETDHNSSSHVLDYQSQKFEKLITN 59  
 ME +YKTLRR+ +NDA I+Q+Q EIF Y ET +N H L Y S EFKI+T+  
 Sbjct: 13 MSFYKTLRKRLINDAHFETIASQALEIFKYGSETYNPKRLHSALGYLSPVFEKIVTH 71

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1938

A DNA sequence (GBSx2047) was identified in *S.agalactiae* <SEQ ID 6009> which encodes the amino acid sequence <SEQ ID 6010>. This protein is predicted to be R5 protein. Analysis of this protein sequence reveals the following:

Possible site: 51  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -3.98 Transmembrane 30 - 46 ( 29 - 51)  
 INTEGRAL Likelihood = -2.76 Transmembrane 967 - 983 ( 966 - 985)

----- Final Results -----  
 bacterial membrane --- Certainty=0.2593(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8935> which encodes amino acid sequence <SEQ ID 8936> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8  
 SRCFLG: 0  
 MoG: Length of UR: 2  
 Peak Value of UR: 2.44  
 Net Charge of CR: 2  
 MoG: Discrim Score: 0.78  
 GvH: Signal Score (-7.5): -0.0599995

-2181-

Possible site: 39  
 >>> Seems to have a cleavable N-term signal seq.  
 Amino Acid Composition: calculated from 40  
 ALOM program count: 0 value: 7.37 threshold: 0.0  
 5 PERIPHERAL Likelihood = 7.37. 194  
 modified ALOM score: -1.97  
 \*\*\* Reasoning Step: 3  
 10 Rule gpol  
 ----- Final Results -----  
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 15 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>  
 LPTKG motif: 944-948

No corresponding DNA sequence was identified in *S.pyogenes*.

20 SEQ ID 8936 (GBS200) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 3; MW 107.4kDa), in Figure 169 (lane 4; MW 122kDa) and in Figure 238 (lane 11; MW 122kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 35 (lane 3; MW 132kDa).

Purified Thio-GBS200-His is shown in Figure 244, lane 9.

25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1939

A DNA sequence (GBSx2048) was identified in *S.agalactiae* <SEQ ID 6011> which encodes the amino acid sequence <SEQ ID 6012>. This protein is predicted to be a 16.1 kDa transcriptional regulator. Analysis  
 30 of this protein sequence reveals the following:

Possible site: 25  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 35 bacterial cytoplasm --- Certainty=0.3919 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9953> which encodes amino acid sequence <SEQ ID 9954>  
 40 was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>CP:CB16108 GB:Z99124 similar to transcriptional regulator (MarR  
 family) [Bacillus subtilis]  
 Identities = 30/114 (26%), Positives = 59/114 (51%), Gaps = 3/114 (2%)  
 45 Query: 29 DVEHLGAPCGHLVWYLYKHPDKMSIKAVEEILHLSKSVASNLVIRMEKNGFIAIVPSKT 88  
 D++ G +LV +Y++P + + + E++ ++ A+ +K++E GFI +P +  
 Sbjct: 25 DDLTLTRGQYLYLVLR-IYENPG--IIQRKLRMIKVDRTTAARAIKLENGQGFICKLDEQ 81  
 50 Query: 89 DKRVKLYLTHLGKKKATQFEIPLKLIHSTLAGITKEEIRITTKKVIRTLARNM 142  
 +K+K L+ T GKK E L+G T EE T ++ + KN+  
 Sbjct: 82 NKKIKKLFFTEKGGKVYFLLRREGEHSTEVALSGPTSEKETISALLHRVRNI 135

-2182-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6013> which encodes the amino acid sequence <SEQ ID 6014>. Analysis of this protein sequence reveals the following:

Possible site: 43  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4175 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 27/64 (42%), Positives = 46/64 (71%)

Query: 3 MENFLQKARILVNLKYLHYAKEXDVEHLAQPGHVLVWLYKPKDKMSIKAVEEILH 62  
 M + R L++Q+E+ D AK+YDVEHLAQPG+++++L KH +++ +K +E+ L  
 Sbjct: 1 MSQVIGDLRELHLHQIEQISDEIAKKQDVEHLAQPGQTVLVFLAKHQHGEIFVKDIEKQLR 60

Query: 63 ISKS 66  
 I +S  
 Sbjct: 61 IFQS 64

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1940

25 A DNA sequence (GBSx2049) was identified in *S.agalactiae* <SEQ ID 6015> which encodes the amino acid sequence <SEQ ID 6016>. This protein is predicted to be 5'-nucleotidase family protein. Analysis of this protein sequence reveals the following:

Possible site: 27  
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -2.66 Transmembrane 668 - 684 ( 665 - 684)

----- Final Results -----  
 bacterial membrane --- Certainty=0.2062 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB12747 GB: Z99108 similar to 5'-nucleotidase [Bacillus subtilis]  
 Identities = 178/535 (33%), Positives = 270/535 (50%), Gaps = 55/535 (10%)

Query: 28 DQGVGVQVIGVNDPFGALDNTGTANMPDQKVNAGSTAQGLD---AYMDQAKDKPKQITPNNG 84  
 + V ++++ +ND HG +D ++ DG GT ++D NY+ + + + K  
 Sbjct: 586 EHVPLRLILGMDNLHGKIDQCYRLDL-DGNQTVDTPTGRMDYAAAYLKEKKAERKN----- 639

Query: 85 ESIRVQAGDMVGASFANSLQLQDEPTVFENFANMNVKYGTILGNHEFDGLAEYNRIVITGKA 144  
 S+ V AGDM+ G S S LLQDEPTV+ + + GT+GNHEFDG E RI+ G  
 Sbjct: 640 -SLIVHAGDMIGSSPVSLLQDEPTVELMDIGFDVGTVGNHEFDGSDLELLRLING-G 697

Query: 145 PAPDSNINNITKSYPHAEAKQRIVVANVIDKVNKQIIPYNNKPYAIKNI PVNKSQVNVGFI 204  
 P +++P ++V AN ++ +P+ +N + V V FI  
 Sbjct: 698 DHPKGTSGYDQGNFF-----LVCANC-----KMKSTGEFFLPAYDIINVBGVVPVAFI 744

Query: 205 GIVTKDIPNLVLRKNYQYEFLEARTIVKYAKELQANQVKAIVLVAHVPAITSKIDIAEG 264  
 G+V++ +V+ + + RF DEA + K A+EL+ K VKAI VLAH+ A + G  
 Sbjct: 745 GVVTSQAAGMVMPEGIKNIEPTDEATVANKAARELKKKGVKIAVLVAHSAEQNGNVAITG 804

Query: 265 EAAEMMKVQNLFPFNSVDIVFAGHNHOYTNGLVGKTRIVQALSQKAYADVRSGLDITDT 324  
 E+A++ K ++ +D++FA HNHQ NG V IVQA GKA V +D T



-2183-

Sbjct: 805 ESADLANKT-----DSRIDVIPAARNHQVNVGKGLIVQAFYKGAIGVVDVIDKTT 859  
 Query: 325 QEFITPSAKVIAVAPGKKTGSADTQAIVDQANTIVKQTEAKIGTAEVSVMITRSVDQD 384  
 +D ++ SA+++V K AI+ + TI + + +G A V + S D D  
 5 Sbjct: 860 KDIVK-KSAEIVYVDQKIEPDSASAILKKYETIAEPIISEVVGDAAVDMRGQYSNDGD 918  
 Query: 385 NVSPVGLSLITRAQLAIARKSWPDIDFAMINNGGIRADILLIKPDGTTITWGAQAQVQPGNI 444  
 +P+G+LI + A + DFA+ N GGIR L G ITWG +QPPGN+  
 10 Sbjct: 919 --TFLGNLITADGNRAAMK-----TDFALMNGGIRAL--KIGSPITWGLYNIQPPGNV 968  
 Query: 445 LQVVEITGRDLYKALNEQYDQKNFFLQIAGLRYYTNDNKEGGEETPFVKVYKAYKSMGEE 504  
 L +EI G+DL +N Q I+G +TT +KE G+ K+ ++G E  
 Sbjct: 969 LTKLEIKGKDLREINAQISFVGPDPYSISG--FTYTDKKTGKAVDMKM----ADGTE 1021  
 15 Query: 505 INFDAKYKLVINDPLFGGGDGFASPRNAKLLGAINP-----DTEVPMAYITDLEK 554  
 I PDA Y L +N+P+ A ++ LIG NP D E +Y+ ++  
 Sbjct: 1022 IQPDATYTLTVNHFMATATG--AKYQFGLGK-NPVTGPEULKATVEYKVSDE 1073

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1607> which encodes the amino acid  
 20 sequence <SEQ ID 1608>. Analysis of this protein sequence reveals the following:

Possible site: 40  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -4.67 Transmembrane 662 - 678 ( 661 - 679)  
 25 INTEGRAL Likelihood = -2.02 Transmembrane 19 - 35 ( 18 - 35)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2869(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 30 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 415/688 (60%), Positives = 517/688 (74%), Gaps = 21/688 (3%)

Query: 1 MKKKIILKSSVVLGLVAGTSMFSSVFADQVGQVIGVNDPFGALDNTGTANMPDGKVA 60  
 35 MKK ILKSSVL ++ +++ +V ADQ VQ +GVNDPFGALDNTGTA P GK+ NA  
 Sbjct: 14 MKKPYILKSSVLSILTSFLLVTDVQADVDVQFLGVNDPFGALDNTGTATTPSKIPNA 73  
 Query: 61 GTAAQLDAYMDDAQKDFKQTNPNGESIRVQAGDMVGASPANSGLIQDEPTVNFNANMVE 120  
 40 GTAAQL AYMDA+ DFKQ N +G SIRVQAGDMVGASPANS LIQDEPTV F N M E  
 Sbjct: 74 GTAAQLGAYMDDAEIDFKQAGDQTSIRVQAGDMVGASPANSALIQDEPTV FVFNKQKPE 133  
 Query: 121 YGTLGNSHFDEGLAEYNRIVTGKAPADPSNINNTKSYPEAAKQSIIVANVIDKNQKI 180  
 YGTLGNSHFDEGL E+NRI+TG+AP P+S IN+ITK Y HEA+ Q IV+ANVIDK K I  
 45 Sbjct: 134 YGTLGNSHFDEGLDFNIRIMTQAPDPSTINDITKQYKRRASHQITVIANVIDKTKYDI 193  
 Query: 181 PYMKPYAIKIPVNNKSVNVGFIGIVTKDIPNLVLRKNYQYEFLOBASTIVKAKELQ 240  
 PY MKPYAIK+I +N+K V +GPIG+VT +IPNLV++N+Y E+FLD AETI KPAKELQ  
 50 Sbjct: 194 PYGKPYAIKDIANDKIVRIGFIVGVTTEIPNLVLKNYHYQFLDVAETIARVAKELQ 253  
 Query: 241 AKNVKAIIVLAHVPAATSKNDIABGAAEAMKKVYNQLFPENSVDIVFAGNHQYTINGLVGK 300  
 ++V AIIVLAHVPAATSK+ + + E A +M+KVNQ++PE+S+DI+FGNHQYTING +GK  
 55 Sbjct: 254 EQHVEAIIVLAHVPAATSKDGVVDHEMATVMEKVNQIYPEHSIDILFAGNHQYTINGIGK 313  
 Query: 301 TRIVQALSQGKAYADVGRGLDITDQFIETPSAKVIAVAPGKKTGSADTQAIVDQANTIV 360  
 TRIVQALSQGKAYADVGR LDTDT DFI+TPSA V+AWAPG KT ++DI+AI++ AN IV  
 60 Sbjct: 314 TRIVQALSQGKAYADVGRGLDITDITDNDFKTPSANVAVAPGKIKTNSD KAINIVAND 373  
 Query: 361 KQVTEAKIGTAEVSVMITRSVDQDNVSPVGLSLITRAQLAIARKSWPDIDFAMINNGGIRA 420  
 K VTE KIGTA S I+++ +D SPVG+L T AQL IA+K++P +DFAMINNGGIR+  
 65 Sbjct: 374 KTVTEKIGTATNSSTISKTEINIDKSPVGNLAIATQLTIAKKTPTTVDFAFNNKNGIS 433  
 Query: 421 DLLIKEDGTTITWGAQAQVQPPGNILQVVEITGRDLYKALNEQYDQKNFFLQIAGLRYYT 480  
 DL++K D TITWGAQAQVQPPGNILQV++TG+ +Y LN+QYD+ Q +FLQ++GL YTY  
 70 Sbjct: 434 DLVENDRTITWGAQAQVQPPGNILQVVIQKTSQRIYDNLQQYDKNQITFLQNSGLTY 493

-2184-

Query: 481 TDNKEGGSETPPKVVKAYKNGREINPDQAKYKLVNDFLGGDGFASPRNAKILGAINP 540  
 TDN +TPPK+VK YK NGBEIN Y +V+NDFL+GGDGF++F+ AKL+GAIN  
 Sbjct: 494 TDNDPKNSDTPPKIVKIVKONGREINLT+TTVVVNDFLYGGDGFSAFKKAKLIGAIN 553

5 Query: 541 DTEVPMAYITDLEKAGKKVSPVNNKPKIYVTMGMVNETITQMDGTHSIIKKLYLDRQNI 600  
 DTE P+ YIT+LE +GK V+ K YVT + + T + G HSII K++ +R GN  
 Sbjct: 554 DTEAFITITNLEASGKTVNATIKGVKNYVTSNLSSTKCVNSAGKHSIIKVPFRNRDQNT 613

10 Query: 601 VAQKSVSDITANQTKSKSTKINPVTTIHKKQLHQFTAINPMRNYGKPSNSTTVSKQLPCT 660  
 V+ B++SD L T++ + + T +N T+ S LP T  
 Sbjct: 614 VSSEVISDLTSTENTNNSLGKKT-----TINKNTLSSTLPIT 653

Query: 661 NSEYQGSFLMSVFG-VGLIGIALNITKK 687  
 Y S +M++ + L G+ KK+  
 15 Sbjct: 654 GDNYYKSPDITLILISLGLNIAFKKR 681

SEQ ID 6016 (GBS328) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 4; MW 73 kDa). The GBS328-His fusion product was purified (Figure 213, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 268), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1941

A DNA sequence (GBSx2050) was identified in *S. agalactiae* <SEQ ID 6017> which encodes the amino acid sequence <SEQ ID 6018>. This protein is predicted to be peptide deformylase (def-2). Analysis of this protein sequence reveals the following:

Possible site: 21  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.70 Transmembrane 55 - 71 ( 55 - 74)

30 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1680 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

35 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB09662 GB:296934 peptide deformylase [Clostridium  
 beijerinckii]  
 Identities = 71/136 (52%), Positives = 96/136 (70%)

40 Query: 1 MIKPIVIRDTFFFLQKQSQMASRADVSLAKDLQFTLHANQNYCVGMAMNIGSLKRVIIINV 60  
 MIKPIV+D FL QKS+ A++ D+ + DL +TL AN +CVG+AA NMIG KR+++ V  
 Sbjct: 1 MIKPIVKDILFLQKQSEATKNDMNVVDDLLDILRANLEHCVGLAANMIGVKRIILVFTV 60

45 Query: 61 GITNLVNMENPVVAKSDPYETERSCLSLVGCSTQRYCHITISYRDINWKEQQLKIDTDP 120  
 G + M NPV++ K PYETERSCLSL+G R T+RY I ++Y D N+ +++ F  
 Sbjct: 61 GNLIVPMINPVILKKEKPYETERSCLSLIGPRKTKRYETIEVTVLDRNPBKKKQVNGFT 120

Query: 121 AQICQHELOHLAGLI 136  
 AQI QHE+DE BGI+I  
 50 Sbjct: 121 AQIQHEMDHFBGIII 136

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6019> which encodes the amino acid sequence <SEQ ID 6020>. Analysis of this protein sequence reveals the following:

Possible site: 45  
 >>> Seems to have no N-terminal signal sequence

-2185-

INTEGRAL Likelihood = -3.61 Transmembrane 55 - 71 ( 55 - 73)

----- Final Results -----

bacterial membrane --- Certainty=0.2444 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 77/136 (56%), Positives = 103/136 (75%)

Query: 1 MIKPIVRDTFPLQKQSOMASRADVSLADLQFTLHANQYCVGMAANMIGSLKRVIIINV 60  
MI+ I+ D F LQKQ+Q+A + D+ +DLQ+TL + C+GMAANMIG KR++I+++

Sbjct: 1 MIRELITDHFLLQKQKQVAKKRDNLIGQDLQDTLAFTRQECGLMAANMIGESKRIVISM 60

Query: 61 GITNLVMFNFPVWVAKSDPYETTESCLSLAGCRSTORYCHITISYREDINWKEQQIKLTDFF 120  
G +LVMFNFP+V+K Y+T+ESCLSL G R TORY IT+ Y D NN+ ++ IT

Sbjct: 61 GPIDLVMFNFPVMVSKRGVYQTKESCLSLGGRKTORYDKITVEYLDHNNRKRKLSLTGLT 120

Query: 121 AQICQHLDHLLSILI 136  
AQICQHLDHLLSILI

Sbjct: 121 AQICQHLDHLLSILI 136

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 25 Example 1942

A DNA sequence (GBSx2051) was identified in *S. agalactiae* <SEQ ID 6021> which encodes the amino acid sequence <SEQ ID 6022>. Analysis of this protein sequence reveals the following:

Possible site: 28  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2880 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA05820 GB:AP001514 NADP-specific glutamate dehydrogenase  
[Bacillus halodurans]

Identities = 298/444 (67%), Positives = 362/444 (81%), Gaps = 2/444 (0%)

Query: 7 YVASVLEKVKKQNEHEESPLQAVEEVFESLVFPVFKYQYIEENLLRLVEPERVISRV 66  
YV V E VK++N +E EF QAV+EVF+SL+PV K+PQY+++ +LER+VEPERVISRV

Sbjct: 16 YVQHVYETVKRRNFWHEFHQAQVKEVFPDSILPVLVKGHPQYVVKQALLERIVEPERVISRV 75

Query: 67 PWDVDDGQVQVNRGVRQPSAIGPYKGLRFPHTVTSIVKPLGPEQIFKNSLTQLTIG 126  
PWDV+G VQVNRG+KVG+R+SA+GPYKGLRFP+V SI+KPLGPEQIFEN+LTG PIG

Sbjct: 76 PWDVDDQVQVNRGVRFPQNSALGPYKGLRFPFVNASIIKPLGPEQIFKNSLTQLTIG 135

Query: 127 GKGKGSNFDPRKGSNDVEMRFTQSPMTELQKYIGPDILVPAGDIGVGRRIQYLYQYKR 186  
GKGKGS+PDPKGS D+MRP+QSPM+EL YIGPD+DVPAGDIGV +RIGY++GQYK+

Sbjct: 136 GKGKGSNFDPRKGSNDGIMRPSQSPMSLSNYIGPDILVPAGDIGVGRRIQYLYQYKR 195

Query: 187 L-NGYQNGVLTKGSLTYGSLARTEATGYGAVFAKMIARAAGQDLTKGVALVSGSGNVA 245  
+ G++ GVLTGSG+ YGSLAR EATGY VVF +EM+ G G +VSGSGNV+

Sbjct: 196 MRGGFPAEVLTKGSLTYGSLARKKATGYGVTVFVRMIKDHGFSFAGSTVTVSGSGNV 255

Query: 246 IYATKEQLGLGATVAVSDSGGVYDFDGLDLETLKQIKVEERARIVKYTEKHFKANFT 305  
IYA EK +LQA VVA SDS GYVD +GIDL+T+K++KEVER RI +Y +HP A++

Sbjct: 256 IYAMEKAMQLGAKVACSDSGGVYDFDGLDLETLKQIKVEERARIVKYTEKHFKANFT 315

Query: 306 ADQGSWSIKADLAFPCATQNKLEDEKALLVENGVLAVTEGAMPSYLGAIKVPQKGV 365

-2186-

G INS+ D+A PCATQNELDE A +L+ NGV AV EGANMPSTL A+ EQ+ GV  
 Sbjct: 316 GCSG-INSVPCDIALPCATQNELDEAAATMLANGVKAWEGBANMPSTLQAVHTFQEHGV 374

Query: 366 AFPGAANAGGVAVSALEMAQNSRRRAWTFEEVDQLRIMKTI PVNASEADEPFGDSG 425  
 F PAKAANAGGVAVSALEMAQNS+R AWTFEKVD +L IMK I+ + +AA+ + SG  
 Sbjct: 375 LFAPAKAANAGGVAVSALEMAQNSTRIAWTFEEVDNGLYRIMKNIYRESIKAEMLYEASG 434

Query: 426 NLVLGANIAGFLKVAQMSAQGIV 449  
 NLV+GANIAGF+KVA AM + G+V  
 Sbjct: 435 NLVVGANIAGFVKVADNKLISGVV 458

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### 15 Example 1943

A DNA sequence (GBSx2052) was identified in *S.agalactiae* <SEQ ID 6023> which encodes the amino acid sequence <SEQ ID 6024>. Analysis of this protein sequence reveals the following:

Possible site: 58  
 >>> Seems to have no N-terminal signal sequence

|    |          |                    |               |                        |
|----|----------|--------------------|---------------|------------------------|
| 20 | INTEGRAL | Likelihood = -8.55 | Transmembrane | 61 - 77 ( 55 - 87)     |
|    | INTEGRAL | Likelihood = -7.70 | Transmembrane | 177 - 193 ( 175 - 202) |
|    | INTEGRAL | Likelihood = -7.06 | Transmembrane | 99 - 115 ( 95 - 122)   |
|    | INTEGRAL | Likelihood = -5.89 | Transmembrane | 42 - 58 ( 40 - 60)     |
| 25 | INTEGRAL | Likelihood = -3.08 | Transmembrane | 160 - 176 ( 159 - 176) |
|    | INTEGRAL | Likelihood = -2.44 | Transmembrane | 124 - 140 ( 122 - 144) |

----- Final Results -----  
 bacterial membrane --- Certainty=0.4418 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9955> which encodes amino acid sequence <SEQ ID 9956> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

### 35 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1944

A DNA sequence (GBSx2053) was identified in *S.agalactiae* <SEQ ID 6025> which encodes the amino acid sequence <SEQ ID 6026>. This protein is predicted to be ABC transporter, ATP-binding protein (msbA). Analysis of this protein sequence reveals the following:

Possible site: 37  
 >>> Seems to have a cleavable N-term signal seq.

|    |          |                     |               |                        |
|----|----------|---------------------|---------------|------------------------|
| 45 | INTEGRAL | Likelihood = -10.72 | Transmembrane | 152 - 168 ( 147 - 192) |
|    | INTEGRAL | Likelihood = -5.47  | Transmembrane | 267 - 283 ( 264 - 288) |
|    | INTEGRAL | Likelihood = -4.30  | Transmembrane | 171 - 187 ( 169 - 192) |
|    | INTEGRAL | Likelihood = -2.13  | Transmembrane | 67 - 83 ( 67 - 83)     |
|    | INTEGRAL | Likelihood = -0.32  | Transmembrane | 493 - 509 ( 493 - 509) |

----- Final Results -----  
 bacterial membrane --- Certainty=0.5288 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 5 >GP: CAB69752 GB: AL137187 putative ABC transporter [Streptomyces coelicolor A3(2)]  
Identities = 269/611 (44%), Positives = 392/611 (64%), Gaps = 31/611 (5%)
- Query: 9 RLWSYLTRYKATLFLAIFLEKLVSSPMSTLEPFILGLAITELTANLV--DMAG----- 59  
RL S +ATLF + V+S ++++ P ILG A + A +V DM G  
10 Sbjct: 27 RLVSQFRPERATLPTLLACVVVSGLNVGPKILGRATOLVFAIGVORIMPSGATKEQVL 86
- Query: 60 -----VSGAEINVPYIAGILIIYFFRGVFFELSGSYGSNYNMITVV 99  
V G ++ + +L++ L + + V  
10 Sbjct: 87 ATMREHODGNVADMLRSTDFVPGQIDFGAVGVBLLJALATFAVAGLLMAVATRLNRAV 146
- Query: 100 QKSIRDRIHDLNRKINKVFSYFDKHQFGUMIGRFTSDVETVSNALQOSPLQIINAFLSI 159  
+++ +R D+ K++++P+SYFDK Q G++L R T+D++ + LQOS Q+IN+ L+I  
15 Sbjct: 147 NRTMFRLREDVQTKLRPLSLSYFDKRGVEVLSRATNDINDIGQTLQOSMQOLINSLLTI 206
- Query: 160 ILVVMVLYLNVLPMIIACIPVTYFSAQAILKRSQPYFKEQAKILGEINFGVQEKLTG 219  
I V+ M+ Y++ LA++ + +P+++ A + KRSQ P +Q + G+LN ++E TG  
20 Sbjct: 207 IGVLMVFYVSMILALVALVTLPVLSFVVATRVGKRSGQFVQCVNRSTGQLNAHIEENTG 266
- Query: 220 FNIIKLYGEEBASSQEFRTIDNLRHVGFKAISFISGIMPFVNLNISDPYILIIAFVQGLQ 279  
++K++GR+E S+++F + D L OFKA F SGIM P++ +S+ Y+++A VQGL+  
25 Sbjct: 267 HAINKVFGRQERSAKQFAEQNDALYEAGFKAQFNSGIMPFVNLNMYLVAVVQGLR 326
- Query: 280 VIAGTLITGRNQAFVQYVWQISQPVQYITQLAGVLQSAKSSLERIFEVL--RESEANQVT 338  
V +G L+IG++QAF+QY Q S P+ + A+ +QS +S ER+FE+LD RE+ A+ +  
30 Sbjct: 327 VASQQLSIGDVQAFIQYSRQFSMFLTVQVSMANLVQSGVASAERVFELLDREQSQADPIP 386
- Query: 339 EKLSDHDTGQVSPFHGVDHFSYDPKPLIRDFNLVDPGQMAIVGPTGAGKTLINILMRP 398  
DL G+V V F Y P+KPLI D +L VRPG +AIVGPTGAGKTLINILMRP  
35 Sbjct: 387 GARPEDLRGRVLEHVSFRYDFEKLPLIRDLGLKVRPGHTVAIVGPTGAGKTLINILMRP 446
- Query: 399 YDVSBGAITVDGHDIRHLSRQDFRQFGMVLDQAMVLYGTIKENLRFG--NLEASDEDIVA 457  
Y+VS G IT+DG DI +SR + R GMVLQD WL+ GTI EN+ +G + E + +I  
40 Sbjct: 447 YEVSGGRITLDGVDIAKMSRDELRAIGTGMVLQDVTWLPGGTIAENTAYGASREVTRGEIE 506
- Query: 458 AAKAANDVDFIRTLPGGYNMVMNQESSNISLGQQLLTIAARLADPKILILDEATSSVD 517  
AA-AA+ D F+RTLP GY+ V++ E + +S G+KQL+TIARA L+DP IL+LDEATSSVD  
45 Sbjct: 507 AARAADHDFVRTLFDGYDTVIDDEGTGVSAGEKQLITIAARLSDPVLILDEATSSVD 566
- Query: 518 TRLELLIQKAMKLMRGRTSFVIAHRLSTIQEADNIIVLKDGQITQENHQKLLADRGFY 577  
TR E+LQKAM KL GRTSFVIAHRLSTI++AD ILV++DG I+EQG H +LL G Y  
50 Sbjct: 567 TRTEVLIQKAMKLAGRTSFVIAHRLSTIRADTTILVMEDGAIVEQGAHTELTADGAY 626
- Query: 578 YELNYSQFSENS 588  
LY +OF+ +  
Sbjct: 627 ARLYKAQARA 637

There is also homology to SEQ IDs 160 and 6546.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1945

- 55 A DNA sequence (GBSx2054) was identified in *S.agalactiae* <SEQ ID 6027> which encodes the amino acid sequence <SEQ ID 6028>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have a cleavable N-term signal seq.

- 60 INTEGRAL Likelihood = -10.88 Transmembrane 242 - 258 ( 235 - 263)  
INTEGRAL Likelihood = -9.82 Transmembrane 159 - 175 ( 129 - 177)  
INTEGRAL Likelihood = -9.71 Transmembrane 52 - 68 ( 49 - 77)

-2188-

INTEGRAL Likelihood = -8.49 Transmembrane 134 - 150 ( 129 - 158)  
 INTEGRAL Likelihood = -1.17 Transmembrane 272 - 288 ( 272 - 289)

----- Final Results -----

bacterial membrane --- Certainty=0.5352 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CA669751 GB:AL137187 putative ABC transporter [Streptomyces  
 coelicolor A3(2)]  
 Identities = 226/565 (40%), Positives = 342/565 (60%), Gaps = 1/565 (0%)

Query: 6 SYLKRYFNWMLDLGLGMLFVTVILSMPTALAGMIDNGVTKDRTGVVYLMFIMFIVVL 65  
 +YL+ Y + L + L L +PT A +ID GV KGD + + +M +  
 Sbjct: 8 TYLRPYKKPIALLVALQFLCTCASLYLPTLNAHIDBGVKGSDGYLSGALMIGISLA 67

Query: 66 GIIGRITAYASSRLTTMIRDMRNDYAKLQYSHIEYBQIGVSSLVTRMTSDTFVLMQ 125  
 ++ I + +R + RD+R ++ ++Q +S E G SL+TR T+D +  
 Sbjct: 68 QVVNIGAVFYGARTAAALGRDVRGAVFDRVQGFARVEVGHGAPSLITRTINDVQQVM 127

Query: 126 FAHSRLRLGLVTPMVMI FSVVMILITSPSLAWLVAVANPLVGVILYVAIKTKPLSERQ 185  
 A M+ L + P++ + +VM L L ++ +P+L + + K+PL + Q  
 Sbjct: 128 LALMTFTLMVASAPIMCVGGIVNALGLDVLPSGVLLGVFVLAICVTLIVKRLRFRIOQ 187

Query: 186 TMLDKINQYVRENITGLRVVRAFARENFPQSCKPQVANQRYTDTSTGLFKLTGLTEPLFVQ 245  
 LD +N+ +RE -TG RV+RAF R+ + +Q+F AN T+ + G L L P+ +  
 Sbjct: 188 VRLDTVNRVLRBQITGNRVIRAFVRDEYRQQRFRKANTSLTEVALGTNLGLMALFRIOQ 247

Query: 246 IIIAMIVAIWVWALDPLQRGAIKIGDLVAFIEYSFHALFSFLLFANLFTNYPRMVSSHR 305  
 ++ +A+VWF + G ++IGDL AF+ Y + S ++ +F M FR V + R  
 Sbjct: 248 VVNLSSIAVWFGAHRIDSGGQIGDLTAFLAYLQIVMSVMATFPMFMVPRVRCAR 307

Query: 306 IREVMMPISINPNTGVTDTKLKGHLEFDNVITFAYGETESPVLHDSIFKAKGETIAF 365  
 I+EV+ + S+ P VT+ + GHLE F YPG E PVL I A+PGET A  
 Sbjct: 308 IQEVLETSSVVPVAPVTHLRKHGHLE+REAGFRYPG-AEEPVLRHIDIVARPGETIAV 366

Query: 366 IGSTGSGKSSLVNLIRPYDVTGLIKILVDGVDRDYNKSLRQKIGFIPQKALLFTGTIG 425  
 IGSTGSGKS+L+ L+PR +D T G++LV+GVDVR + K+L + + +PQK LF GT+  
 Sbjct: 367 IGSTGSGKSTLLGLVRLFDATIDGTVIVGVVDVTRVDEKFLAKVSVLPQKPYLFAQVA 426

Query: 426 ENLKYGKADATIDDLRQAVDISCAKEFIESHQEAFETHLAEQSGNLSGGQKQRLSTARAV 485  
 NL+YG DAT ++L A+ ++QAKET+ + + +A+QG+N+SGGQ+QRL+IAR +  
 Sbjct: 427 TNLRYGNPDATDEELMHALAVACAKEFVSEGLEGLDAPIAQGGTNVSGGQRLAIARTH 486

Query: 486 VKDPDLYIFDSSSALDYKTDNTAIRLKEVTGDSVTLIVAQRVGTIMADQIIVLDRGE 545  
 V+ P++Y+FDSSSALDY TDA LRA L + T ++TV+IVAQRV TI DAD+I+VLDEG  
 Sbjct: 487 VQRPEYILFDDSSSALDYATDAALRAELAQSTAEATVIVIAQRVATIRADRIIVLDEGR 546

Query: 546 IVRGTHTAQLIENNAIYREIAESQL 570  
 +VG GH +L+ +M YREI SQL  
 Sbjct: 547 VVGVGRIHELMADNETHREIVLSQL 571

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4985> which encodes the amino acid  
 sequence <SEQ ID 4986>. Analysis of this protein sequence reveals the following:

Possible site: 22

&gt;&gt;&gt; Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -16.24 Transmembrane 155 - 171 ( 145 - 176)  
 INTEGRAL Likelihood = -7.48 Transmembrane 130 - 146 ( 122 - 150)  
 INTEGRAL Likelihood = -5.04 Transmembrane 13 - 29 ( 12 - 30)  
 INTEGRAL Likelihood = -5.04 Transmembrane 56 - 72 ( 52 - 75)  
 INTEGRAL Likelihood = -4.14 Transmembrane 239 - 255 ( 238 - 259)  
 INTEGRAL Likelihood = -1.70 Transmembrane 269 - 285 ( 269 - 288)

----- Final Results -----

-2189-

bacterial membrane --- Certainty=0.7496 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 175/511 (34%), Positives = 296/511 (57%), Gaps = 3/511 (0%)

Query: 59 MFPIFVVLGIIGRITMAYASSRLTITMIRDMRNDMYAKLQEQSHHEYEQIGVSSLVTRMTS 118  
 + I + LG++ + + + + + DNR + K+Q++S+ E +LV R+T+  
 10 Sbjct: 56 LIIALLGLMMSGAINTVLAAKIAGQGVSDMRREKTFRKIQDPSYANIEAPNAGNLVRLTN 115

Query: 119 DTFVIMQFARMSLRGLVTPMVMI FSVVMILITSPSLAWLVAVAMPLLGVILYVAIKTK 178  
 D + M ++ P++ I + +M + T P L W++ V + L+ ++ V +  
 15 Sbjct: 116 DINIQSLVMVMFQILFRLPILFIGA FIMAVQTFPQMWVIVVMVILIALMGLVMRQMG 175

Query: 179 PLSEKQQTMDKINQYVRENLTGLRVVRAFARENFSQKQFVANQRYTDTSTGLFKLTGL 238  
 P + Q ++D(KIN+ +ENL G+RVV++F +E Q KF+ + + + L  
 20 Sbjct: 176 PRFGKQRLMD(KINRAKENLRGVVVKSFVQBQQQYTKFKETSDNLLALNLSIGYGFSL 235

Query: 239 TEPLFVQIILAMIVAIVFALDPLQRAIKIGDLVAFIEYSFHALFSPLFANLFTMYPR 298  
 +P + + + + ++ IG++ +F+ Y +FS ++ ++ R  
 25 Sbjct: 236 MQPALMLVSYLAVTVSINVVTVEIDPTVIGNIASFTMYMQIMFSIIVGSGMGNQVSR 295

Query: 299 MVSRRIRREVMNPISINPNTGVTIDTKLKGHLFDMVTFAYGETESPVLRHDSFKAK 358  
 VS RIR+++ ++ E + + G + PD+V+F YP + E P L ISF +  
 30 Sbjct: 296 APVSGARIRQLISLTPAMTFENE--KEETISGSIVFDVVSFTYNDDE-PTLKHISFAIE 352

Query: 359 PGETIAFIGSTGSKSSLVNLI PRFYDVTIGKILVDGVDVRDYNLKSIRQKIGFIPQKAL 418  
 PG+ + +G+TSSGKS+L LIPR +D G+IL+ G ++ +LRQ + + QKA+  
 35 Sbjct: 353 PGQVGVGVGATSGSKSTLAQLIPRLFPDQDQLLGGPKIKTLGQTTLRGSVILVIGKI 412

Query: 419 LFTGTITGENLKYGKADATIDDLRQAVDISQAKEFTBSHCFAFETHLAEQSGNLSGGQKQR 478  
 LF+GTI +HL+ G A A ID + + + + I+QAKEFI+ +E+ +E GSNLSGGQKQR  
 40 Sbjct: 413 LFSGTITADNLQSGSAKADIDAMQKAAIQAKKEFTIDRMDRYSQSVEERGSGNLSGGQKQR 472

Query: 479 LSIARAVVKDPOLYI FDDSFSAIDYKTDATLARLKEVIGDSTVLIVAQRVGTIMADQI 538  
 LSIAR V+ P + I DDS SALD K++ ++ L +T +IVAQ++ ++ AD+I  
 45 Sbjct: 473 LSIARGVINHPKILILDOSTSALDAKSEKRVQALSHKLGTTTIVIAQRISSVKADKI 532

Query: 539 IVLDBGEIVGRGTHAQLIENNAIYREIAESQ 569  
 +VLD+G+++G GTHA+L+ NNAIYREI E+Q  
 50 Sbjct: 533 LVLDQQQLIGEGTHAELVANNAIYREIYETQ 563

There is also homology to SEQ IDs 72 and 6552.

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1946

A DNA sequence (GBSx2055) was identified in *S.agalactiae* <SEQ ID 6029> which encodes the amino acid sequence <SEQ ID 6030>. Analysis of this protein sequence reveals the following:

50 Possible site: 24  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.2391 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA51784 GB:X73368 ORF 18.3 [Salmonella typhimurium]

-2190-

Identities = 58/162 (35%), Positives = 92/162 (55%), Gaps = 8/162 (4%)

Query: 1 MIIRPIIKNDQQAVALRQSLRAYDL--DKPTATSDPHLDHLTSYEEKIEKSGFVIR 58  
 + +R I + A+A+IRQ Y L DK T +DP+LD L Y + + ++V+E  
 Sbjct: 9 LTVRRITADNAAIARIQVSAEYGLTADKGYTV-ADPNLDLYQVYSQ-PGAAYVWVE 66

Query: 59 ERDEIIIGGGFGPLKHL---IARMQKVYIAERPRGKGLATDLVKMIEVEARKIGYRQLYL 115  
 + ++G GG PL I R+QK+Y RG+GLA L M AR+ G+++ YL  
 Sbjct: 67 QNGCVVGGGSGVAPLSCSEPDICLRQMYFLPVRIGQGLAKKIALMALDHAREQGFRCYLL 126

Query: 116 ETASTLSRATAVYKIMGYCALSQFIANDQGHETAMDIWMIKDL 157  
 ET + L A A+Y+ +G+ +S+P+ GH ++ M+KDL  
 Sbjct: 127 ETTAFLREATAIYERIGFPHISEPL-GCTGHVDCEVRMLKDL 167

- 15 No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1947

- 20 A DNA sequence (GBSx2056) was identified in *S. agalactiae* <SEQ ID 6031> which encodes the amino acid sequence <SEQ ID 6032>. This protein is predicted to be ABC transporter. Analysis of this protein sequence reveals the following:

Possible site: 25  
 >>> Seems to have no N-terminal signal sequence

- 25 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1738 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 30 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12566 GB:299108 similar to ABC transporter (ATP-binding  
 protein) [Bacillus subtilis]  
 Identities = 269/625 (43%), Positives = 397/625 (63%), Gaps = 11/625 (1%)

- 35 Query: 1 MSDFLVDGLTKSVGDKTVFSNVFSFIHSLDRIGIIGVNGTKPTLLDVISGELGFGHRS 60  
 MS + L K+ GDKT+F ++SF I +RIG+IG NGTK+TLL VI+G +  
 Sbjct: 1 MSILKAENLYKTYGDKTLFDHISPHIRENERIGLIGPNTGKSTLLKVIAGLESIE--EG 58
- Query: 61 PFSSANDYKIAIYKQZEPFDDSQITLDTVLSSDLREMAIKYELLINHY-----EBSQ 115  
 + + ++ L Q+P+ Q+L+ + S + M ++EYE L E +Q  
 40 Sbjct: 59 EITKSGSVQVEFLHQDPELPAQTVLHRTYSGESAVMKTLEREKALYELGKDPEHRCQ 118
- Query: 116 SRLRKVMABMDSLDWASIESEVTKVSLKLGITDLQLSVGELSGELRRRVQLAQLVLIAD 175  
 L A+MD+ +W + KTVLSKLG+ D+ V ELSSG ++RV +A+ L+ AD  
 45 Sbjct: 119 KHLAAQANDANNANDANTLAKTVLSKLGVDVTKPVNLSGSGQKKRVATAKMLQIAD 178
- Query: 176 LLLLEDEPTNHLIDITIAWLINFLNSKICVLFITHDRYFLONVATRI FELDKAQITSYQS 235  
 LL+LDEPTNHLID +TI WL +L V+ +THDRYFL+ V RI+EL++ + Y+G  
 Sbjct: 179 LLLLEDEPTNHLINSETIENGLYSQPGAVMLVTHDRYFLANRVNIRIYELRGSGLTYGK 238
- 50 Query: 236 NYQDYVIRKARQDRDAASLHKKKQLYKQELAWMTQPOARATKQQAIRNFQNLKDLH 295  
 NY+ ++ RAE++ + K+ L +ELAW+R +AR+TKQ+ARI+R + LK  
 Sbjct: 239 NYEVLKRAEREAQAEQKSTKRNILRLRELAWLRGAKARSTQKARIDRVETLKEQTG 298
- 55 Query: 296 QTSITSDLEMTFETSRIGKKVINPKNVSPSPYOKSILKDPNLLIQKDRIGIVGNGVWK 355  
 S S L+ + R+GK+VI ENV +Y + + FN L+ +RIGI+G NG+GK  
 Sbjct: 299 PQSSGS-LDPAIGSHRLGQVTEAENVMIAYDGRMLVDRFNELVIGPERIGIIGPGRIGK 357
- 60 Query: 356 STLLNLIVQDLQPDSCGNVSGIETIRVGYFSQQLHMDGSKRVINYIQLQEVADSVKTSVGT 415  
 +TLLN + ED G+++IG+T+R+GY++Q M+G +VI+Y+E A+ VRT+ G



-2191-

Sbjct: 358 TLLNALAGRHTPDGGDITIGQTVRIGYITQDHSBNGELKVIDYIKETAIEVVKTADGDM 417

Query: 416 SVTE-LLBQLFPRSTHGTQIAKLSGGRKKRLYLLKILIEKPNVLLDEPTNDLDIATLT 474  
E +LE+FLFPRS T I KLSGGRK+RLYLL+L+++PNVL LDEPTNDLD TL+

5 Sbjct: 418 ITABQMRLERFLFPRSQQTYIRKLSGGRKKRLYLLQVLMQRPVVLFLDEPTNDLDITETLS 477

Query: 475 VLENFLQGGFGGVITVSHDRYFLDKVANKIIAIFEND-IRBFGNYTDYLDKAFNEQNN 533  
VLE+++ F G VITVSHDRYFLD+V ++I FE N I F G++DY++E +

10 Sbjct: 478 VLEDYIDQFGVGVITVSHDRYFLDRVDRLLVFEENGVISRFQGSYDYDMERSKAKKAAP 537

Query: 534 EVISKKESTKTSRRKQSKRMSYFRKQRMATIEDDILENTITIRINMDTCGSDFTRL 593  
+ + +E T + K+ RK++SY ++ HW IED I LE ++E D+ GSDF ++

15 Sbjct: 538 KP-AAEKTARAEAPKKKKLKYSDQLEMDGIEDKIAQLEKHKHQLADIAAAGSDFGKI 596

Query: 594 SDLOKELDAKNEALLEKDYDRYEYLS 618  
+L E E L DR+ LS

Sbjct: 597 QSLMAEQAKTAELEAAMDRTWELS 621

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6033> which encodes the amino acid  
20 sequence <SEQ ID 6034>. Analysis of this protein sequence reveals the following:

Possible site: 60  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
25 bacterial cytoplasm --- Certainty=0.2591 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 467/624 (74%), Positives = 535/624 (84%), Gaps = 3/624 (0%)

Query: 1 MSDFLVDGLTKSVGDKTVFSNVSPFIHSLDRIGIIGVNGTGKTTLLDVISGELGFGDGRS 60  
MS FLV+ LTK+VGDKTVF ++SFIH DRIGIIGVNGTGKTTLLD+V+SG LGFGD S .

35 Sbjct: 1 MSHFLVEKLTKTVGDKTVQFDISFIHDFDRIGIIGVNGTGKTTLLDVLGSRGLGFGDGRS 60

Query: 61 PFSSANDYKIAIYLKQEPDQDSDQITILDVLSDDLRENALIKYEYELLNHYEBSKQSRLEK 120  
PFS ANDYKIAIYL Q++F+D+ ++LDTVLS+D++ + Li++YELL+ +Y E KQ LE

Sbjct: 61 PFSSANDYKIAIYLTQDFPNDASVLTDLVLSADVKAATQLIRQTELLMANYTDEKQSRLEK 120

40 Query: 121 VWAEXDSDLDWAGSIBSEVTKVLSKLGITDQLSLVGLSGGLRRRVQAQLVLLNDADLLLLL 180  
+M+EMD LDWAGSIB+VTKVLSKLGITDL+ VG+LSGG+RRRVQAQLVLL ADLLLLL

Sbjct: 121 LMSEVRLDLDWAGSIBSVTKVLSKLGITDQLQVGLSGGMRFRVQAQLVLLGADLLLLL 180

45 Query: 181 EPTNHLDDITIAWLTINFLANSKTKVLFIETHDRYFLDMVATRI FELDKAQITEYQGNQYDY 240  
EPTNHLDDITIAWLT +LK +KTKVLFIETHDRYFLD+VATRI FELDKA +TEYQGNQYDY

Sbjct: 181 EPTNHLDDITIAWLTITLTKAKTKVLFIETHDRYFLDMVATRI FELDKAGTKLEYQGNQYDY 240

Query: 241 VRLAQQDERDAASLHKKKQLYKQELAWMTQPPQARATQQAIRNFPQNLKNDLHGTSDT 300  
VRL+AQDERDAA+LHKKKQLYKQELAWMTQPPQARATQQAIRNFP +LK ++HQ E

50 Sbjct: 241 VRLKAQDERDAANLHKKKQLYKQELAWMTQPPQARATQQAIRNFPSDLKKEVGDSSA 300

Query: 301 SDLEMTFSTRIGKKVINFENVSFSPYSKILKDFNLLIQNKDRIGIVGDNGVKGSTLLN 360  
LEMTFSTRIGKKVI+FE++SP+Y D+ ++KDFNL+IQNKDRIGIVGDNGVKGSTLLN

Sbjct: 301 DKLEMTFSTRIGKKVITHFEDLSFAYGDRQLIKDFNLIQNKDRIGIVGDNGVKGSTLLN 360

55 Query: 361 LIVDQLPDGNGVISTETIRVGYFSQQLANMDGSKRVINFLQVADAEKTSVGTTSVTEL 420  
+I DL+P SG + IG+TIRVGYFSQQL +MD +KRVINFLQVADAEKTSVGTTS++EL

Sbjct: 361 IINGDLKPTSGKLDIGUTIRVGYFSQQLKMDDEKRVINFLQVADAEKTSVGTTSISEL 420

60 Query: 421 LBQFLFPRSTHGTQIAKLSGGRKKRLYLLKILIEKPNVLLDEPTNDLDIATLTVLENFL 480  
LBQFLFPRS+HGT IAKLSGGRKKRLYLLK+LIEKPNVLLDEPTNDLDIATLTVLENFL

Sbjct: 421 LBQFLFPRSHGTTLIAKLSGGRKKRLYLLKILIEKPNVLLDEPTNDLDIATLTVLENFL 480

Query: 481 QGFGGSPVITVSHDRYFLDKVANKIIAIFENDIRBFGNYTDYLDKAFNEQNNVISKKE 540  
F GPVITVSHDRYFLDKVA KI+AFE+ DIR P+QNY+DYLDK F ++ E K

65

-2192-

Sbjct: 481 ANFAGPVITVSHDRYFLOKVATKILAFEEGDIVFYGNYSYDLDEKVFKEKTVADLAKT 540

Query: 541 STKTS---REKQSRKRMGYFEKQEWATIEDDIMLENTYTRIENMQTCGSDPTRLSDIQ 597  
+ +K+ KRRMSY EKQEWALIED I +E I IEN M T SD+ +L+ LQ

5 Sbjct: 541 TVTREVPLPQREERKMSYLEKQEWATIEDKIATIRANIEIRNQMIVVSDYDGLQLQ 600

Query: 598 KELDAKNEALLEKDYRYEYLSLD 621  
KELD +N LL Y+R+EYLS LD

10 Sbjct: 601 KELDQRNDLLLAYERPEYLSGLD 624

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1948

A DNA sequence (GBSx2057) was identified in *S. agalactiae* <SEQ ID 6035> which encodes the amino acid sequence <SEQ ID 6036>. This protein is predicted to be poly(a) polymerase (papS). Analysis of this protein sequence reveals the following:

Possible site: 14  
>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.2658 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

25 A related GBS nucleic acid sequence <SEQ ID 9957> which encodes amino acid sequence <SEQ ID 9958> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA838446 GB:147709 poly(A) polymerase [Bacillus subtilis]  
Identities = 157/395 (39%), Positives = 235/395 (58%), Gaps = 14/395 (3%)

30 Query: 11 FQKALPILKKIKAGYEAYFVGGSVRDVLDRPIHDVDIATSSYPESTKQIPKRTVDVGI 70  
F KALP+L+ +AG++AYFVG+VRD + R I DVDIAT + P++ +++F+RTVDVG  
Sbjct: 5 FIKALPVLRLIEAGHCAYFVGGAVRDSYMKRTTIGDVDIATDAAPDQVERLPQRTVDGK 64

35 Query: 71 EHGTVLVLKGGYEYITTFRTVEYVYDVRFPQVNFVRSLEEDLKRRDPTNAPALNEDG 130  
EHGT++VL + YE+TTFRTE YVD+RRFS+V F+ SLEEDLKRRD T+N A+ DG  
Sbjct: 65 EHGTVILWDEETYEYITTFRTSDYVDFRFPSEVQFISSLEEDLKRRDLTINAWMTADG 124

40 Query: 131 EVIDLPHGLDLDLNLRLRAVGLASERFNEALRIMGLRFPASLNFDIETTTTFEAMKHA 190  
+V+D F G D+D ++R VG +RF EDALR++R +RF + L F + T RA+ K  
Sbjct: 125 KVLIDYFGGKKIDQKIVRTVQKPEDRFQEDALRMLRAVRFSQGLGFTLSPETERALAEK 184

45 Query: 191 SLLEKISVERSFIEFQKLLAPYWRKGMALALIDSHAFNYLPCLKNRELQLSAFLSOLDKD 250  
SLL ++SVR IEF+KLL R+ + LI + + LP ++ L +S +  
Sbjct: 185 SLSSHVSVERKRTIEFKLQGRASRQALQTLQTRLYEELPGFYHKRETL---ISTEFP 241

50 Query: 251 FLFETS-EQAWASLILSMEV--EHTKTLKKWKISTHFQKDVESHIVDVYRIRQMGILTKS 307  
F TS R+ WA+L+ +++ + + FLK WK K+ HI D + + L  
Sbjct: 242 FFSLSIREELWALLINGLVILKQAPLFLKAWKLPKGVKICAHIAITF----QGSIDAM 297

55 Query: 308 HLYRYGKTIKQAGIRKAR-GIMVDFEKIEQLD---SKIAIHDRHRIVUNGKDTIILKIG 363  
+YR GK + A I + R +D +K++ + L I ++ + G L+  
Sbjct: 298 TMYRAGKALLSAKISQLRQNEKLEKKIKDIOYAYQNLPIKSLKLDITGKDLALRN 357

Query: 364 IKPQPQMGDIISQIELAIVLGLQNLNEERAILHFVK 398  
G + + + IE A+V G+L N+++ I ++K  
Sbjct: 358 RPAGKQVSEELQWIBQAVVTGKLSNQKKHIEEWLK 392

-2193-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6037> which encodes the amino acid sequence <SEQ ID 6038>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2023 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 256/400 (64%), Positives = 312/400 (78%)

```

Query: 2 MRINYLPSFQKALFILKKIKKAYFVGGSVRDVLLRPIHVDVIATSSYPETKQI 61
 M-L +PSEFQKALFIL KIK+AGYKAYFVGGSVRDVLL+RPIHVDVIATSSYPETK I
Sbjct: 1 MKLMTMPSEFQKALFILTKIKEAGYKAYFVGGSVRDVLLRPIHVDVIATSSYPETKAI 60

Query: 62 FKKITVDVGIEHGTVLVLEKGGYEITTFRTIEVYVYRRPSQVNFVRSLEEDLKRRDFTV 121
 F RITVDVGIEHGTVLVLE GGEYEITTFRTIE++YVDYRRPSQV+ FVRSLEEDLKRRDFTV
Sbjct: 61 FNKITVDVGIEHGTVLVLENGGGYEITTFRTIEDIYVDYRRPSQVSFVRSLEEDLKRRDFTV 120

Query: 122 NAFALNEDGEVIDLDFHGLDLDNHLRAVGLASERFNEDALRIMRGLRFSASINFDIETT 181
 NA AL+S+G+VID F GL DL LRAVG A ERF EDALRIMRG RP+ASL+FDIE
Sbjct: 121 NALALDENGQVIDKFRGLDLDKQKRLRAVGLASERFEEDALRIMRGFRFASLDFDIEAI 180

Query: 182 TFEAMKKHASLLEKISVERSFIEDKLLAPYWRKGMALIDSHAFNYLFLCNRELQLS 241
 TFEAM+ H+ LLEKISVERSF EFDKLL+AP+WRK+ A+I A++YLP LK +E L+
Sbjct: 181 TFEAMRSHSLLEKISVERSFTFEDKLLMAPHWRKGISAMIAQQAYDYLGLKQKEAGLN 240

Query: 242 AFLSQLDKDLPFETSEQWASLILSNEVHTKTFLLKWKTSSTHFKQDVEHIVDVIYRIR 301
 + L + F F QWAA +++S+ +E K+FLK WKTS FQ+ V ++ +YRIR++
Sbjct: 241 HLIVSLKDNFTSDYHQWAWYVMSLAIEDPKSFLGAWKTSNDQRYVTKLIALYRIQE 300

Query: 302 MGLTKEHLRYRGKTIKQAGIRKARGMLVDFEKIQDLSLAIEDHREHIVNGGTLIKK 361
 K +Y+YKG + E +RKA+ L VD ++I LD L IEDH+IV+NG LIK
Sbjct: 301 RSFEKLDIYQYGGKMASLVEDLRKAQSLSVMDRINTLDQALYIHDSHIVLNGSLIND 360

Query: 362 LGIKPGPQWGDIIISQIBLAIVLQGLNEEAILHFVKOYL 401
 G+K GPQ+G ++ +ELAIV G+L N+ I FV++ L
Sbjct: 361 FGWKGSGPQLGLMLEKVELAIVGRILDNDFTTIEAFVREEL 400

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1949

45 A DNA sequence (GBSx2058) was identified in *S.agalactiae* <SEQ ID 6039> which encodes the amino acid sequence <SEQ ID 6040>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2939 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

55 The protein has homology with the following sequences in the GENPEPT database.

>GP:BA007346 GB:AP001519 unknown conserved protein [Bacillus halodurans]  
Identities = 94/274 (34%), Positives = 153/274 (55%), Gaps = 2/274 (0%)

Query: 2 KIALITDTSAYLFAIENIHEDVYLDIPIIIDSKTYTSGQNLTLDQYDKLAASKELPKT 61

-2194-

K+A+TD++AYL V V+ + ++ + Y E L+ +Y+KL ++LP T  
 Sbjct: 3 KIAIVTDSATYALGPKRAKELGVIVVPLSVVGERAYQESVELSSADFTKLAGHEEKLPIT 62  
 Query: 62 SQPSLAELDDLQLEKEBGYTHVLGLFIAAGISGFQWNIQFLIEHPNLTIAFPDPTKITS 121  
 5 SQP++ + +L KBG+ V+ + +++ ISG +Q+ + + D+ I+  
 Sbjct: 63 SQPAVGLFVETTERLAKBGFVVI SIHLSSKISGTQSGALTAGSMVGGIVIGYDSGISG 122  
 Query: 122 APQGNLVRNALMCSRGMDFDVIVNKIQSIEKIEGFIIVNDLNLHVKGRILNSGSAIG 181  
 PQ N V A +EG D I++ ++ VV+DL+HL +GRL+ ++G  
 10 Sbjct: 123 EPQNFVARAAKLVKEGADPQTIIDHLDEVKRRTNALFVVDLSHHRGGRINAAQLVVG 182  
 Query: 182 NLLSIKPVLFHNEEGKIVVYKVRTEKKALRIARI -VKEMTADGEYDIAIHSRAQDKA 240  
 +LL IKP+LHF E+G IV EKVTEKKA R+ E+ +E ++ +IH+ D A  
 Sbjct: 183 SLLKIKPILHF -EDGSIVPLEKVRTEKKAMRVKELFAEASSASSVKAIVIHANRLDGA 241  
 Query: 241 EQYNLLAKAGLKDDLEIVSGGVIATHLGEBAV 274  
 E+L + + D+ I PG VI THLGE++  
 Sbjct: 242 EKLADERSQSFHVVDVISRHPGPIVIGTHLGEBSI 275

- 20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6041> which encodes the amino acid sequence <SEQ ID 6042>. Analysis of this protein sequence reveals the following:

Possible site: 14  
 >>> Seems to have no N-terminal signal sequence

- 25 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3379 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 30 An alignment of the GAS and GBS proteins is shown below.

Identities = 161/281 (64%), Positives = 233/281 (82%)

Query: 1 MKLALITDTSAYLFRATENHRDVIYLDIPIIDDKTYIEBQNLITLDQYVYDKLAASKELPK 60  
 35 Sbjct: 1 MKLAVITDSTATLPTDLKQDAIFSLDIPVITDDIETPEGRNLISIDDFYQNRADSQNLPK 60  
 Query: 61 TSQPSLAELDDLQLEKEBGYTHVLGLFIAAGISGFQWNIQFLIEHPNLTIAFPDPTKIT 120  
 TSQPSL+ELD+LL L +GYTHV+GLF+A GISGFQWNIQFL EHP + +AFPD+KIT  
 40 Sbjct: 61 TSQPSLELDNLGLLSKGYTHVIGLFLAGISGFQWNIQFLAEHPEIEMAFPDKIT 120  
 Query: 121 SAPQGNLVRNALMCSRGMDFDVIVNKIQSIEKIEGFIIVNDLNLHVKGRILNSGSAII 180  
 SAP G++V+N L SR+GM F I+NK+Q QI+ FI+V+DLNLHVKGRILNSGSA++  
 Sbjct: 121 SAPLGSMVNRVLDSRQGMTFQAILNLKLEQIDGTATFIMVDDLNLHVKGRILNSGALL 180  
 Query: 181 GNLLSIKPVLFHNEEGKIVVYKVRTEKKALRIARIVKMTADGEYDIAIHSRAQDKA 240  
 45 GNLLSIKP+L F+EEGKIVVYKVRTEKKA+KRL EI+ ++ ADG+Y++ IHS+AQDKA  
 Sbjct: 181 GNLLSIKPIILRFDEEGKIVVYKVRTEKKAMKRLVELLNDLIADGQYINVFITHSKAQDKA 240  
 Query: 241 EQYNLLAKAGLKDDLEIVSGGVIATHLGEBAVAGITPK 281  
 + L LL +G + D+E V FG VIATHLGEBA+AFG+TF+  
 50 Sbjct: 241 DYLKRLQDSQYQYDESVHGRVIAITHLGEBAIAGVTPR 281

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 55 Example 1950

A DNA sequence (GBSx2059) was identified in *S.agalactiae* <SEQ ID 6043> which encodes the amino acid sequence <SEQ ID 6044>. Analysis of this protein sequence reveals the following:

Possible site: 46  
 >>> Seems to have no N-terminal signal sequence  
 60 INTEGRAL Likelihood = -1.59 Transmembrane 51 - 67 ( 50 - 67)

-2195-

## ----- Final Results -----

bacterial membrane --- Certainty=0.1638 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6045> which encodes the amino acid sequence <SEQ ID 6046>. Analysis of this protein sequence reveals the following:

Possible site: 61

>>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -3.19 Transmembrane 50 - 66 ( 49 - 67)

## ----- Final Results -----

bacterial membrane --- Certainty=0.2275 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

Identities = 94/126 (74%), Positives = 115/126 (90%)

Query: 1 MEVIRBQEFVNQHYDARNLEWHEENGTPKINFEVTFQLANRDEAAKVTISIAVLQFVIV 60  
 M+++RB+EFVNQHYDARNLEW+KNGTP+TNFEVTFQL ++DE K T IV+VLQFVIV  
 Sbjct: 1 MQLVRKEEFVNQHYDARNLEWHEENGTPETNFEVTFQLIDKDBQQKBTIVSVISVLQFVIV 60

Query: 61 RDEFVISGVISQMAHIQGRLINEPSEFSDQEVENLAAPLEIVKRLTYEVTEIALDRPGV 120  
 ++EFVISGVISQM I RL+++PSEF+Q+EVE+LAAPLL++VIGRLTYEVTEIALDRPG+  
 Sbjct: 61 KEEFVISGVISQWRIIDRLVNDKPEPTQEEVESLAAPLLDMVIGRLTYEVTEIALDRPGI 120

Query: 121 TLEFNS 126  
 LEF +  
 Sbjct: 121 HLEFKN 126

SEQ ID 6044 (GBS416) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 4; MW 17.5kDa).

GBS416-His was purified as shown in Figure 214, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 1951

A DNA sequence (GBSx2060) was identified in *S.agalactiae* <SEQ ID 6047> which encodes the amino acid sequence <SEQ ID 6048>. Analysis of this protein sequence reveals the following:

Possible site: 36

>>> Seems to have no N-terminal signal sequence

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3875 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-2196-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1952

- A DNA sequence (GBSx2061) was identified in *S.agalactiae* <SEQ ID 6049> which encodes the amino acid sequence <SEQ ID 6050>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

- No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1953

- A DNA sequence (GBSx2062) was identified in *S.agalactiae* <SEQ ID 6051> which encodes the amino acid sequence <SEQ ID 6052>. This protein is predicted to be PTS system, fructose-specific enzyme II, BC component (fruA-1). Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

25 INTEGRAL Likelihood = -10.56 Transmembrane 630 - 646 (618 - 653)
INTEGRAL Likelihood = -7.43 Transmembrane 307 - 323 (303 - 331)
INTEGRAL Likelihood = -7.01 Transmembrane 415 - 431 (412 - 435)
INTEGRAL Likelihood = -7.01 Transmembrane 448 - 464 (444 - 474)
INTEGRAL Likelihood = -3.72 Transmembrane 595 - 611 (591 - 612)
INTEGRAL Likelihood = -3.61 Transmembrane 530 - 546 (529 - 553)
30 INTEGRAL Likelihood = -2.39 Transmembrane 350 - 366 (350 - 371)
INTEGRAL Likelihood = -1.70 Transmembrane 486 - 502 (486 - 506)
INTEGRAL Likelihood = -1.49 Transmembrane 376 - 392 (376 - 392)

----- Final Results -----
35 bacterial membrane --- Certainty=0.5225 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

- A related GBS nucleic acid sequence <SEQ ID 9959> which encodes amino acid sequence <SEQ ID 9960> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

- ```
>GP:BAB04547 GB:AP001510 PTS system, fructose-specific enzyme II, BC
component [Bacillus halodurans]
Identities = 320/659 (48%), Positives = 438/659 (65%), Gaps = 46/659 (6%)

45 Query: 1 MKIQDLLKKEVMIMDLKATSKBAIDEMITKLVDTGVVTFNFAIFKDGIMGREAQTSGLG 60
+KI +LLKK+ M++I+I+A SKRA IDE++ L G + + FK I++RE++Q+TG+G
Sbjct: 2 LKISELLKKDITVNLNRAASKAVIDELVRLTDKAGRLNDRAQFKRAILERSQSTTGVG 61

50 Query: 61 DGIAMPHSGKNAAVKRAIVLFAKASGVYVYALDQPTDLFFNIAAPDGNATHLAALAE 120
+GIA+PH+K AAVK+ + F +S +G+UVE+LDQP+ LFFNIAA +GAM+ HL L+ L
Sbjct: 62 EGIAIPHAKTAAVKQFALAFGRSDNGIDYSELQDQPSHLFFNIAAEGANNEHLETSL 121
```

-2197-

Query: 121 SKYLLKRGFADQLRQAKFPDIIATFDSNISISQRTVAPQTVQSTSGSDYIVACTTIG 180
S +L+ E F L +A++ D+I+A D +E + +G + ++AVT C TG
Sbjct: 122 STFLMDETPRSTLMAKQSEDEILAALD---KKEARTAGEAEKQESYE-LLAIVTGCPTG 176

Query: 181 IAHTYMAEALKKKAEMSGVGIKVTNGSAGVGKNTSSDIARAKGVIIAADKAVEMDRP 240
IAHTYMA + LK KA E+GV IKVSTNG+ GV N+LT +I+ AK +I+AAD VEMDRP
Sbjct: 177 IAHTYMAADNLKSKAGBLGVSIVKVTNGSGGVKNRLTDEELSAKALIVAAIDTVKEMDRF 236

Query: 241 DGRFLVSRPVADGKKKSRLINILNKAQTYHAKNQNDKQSGESDKSGLSG---AFYK 297
GRF++ PV DGI++ ++LI+ L KA Y + Q+ DG +G G FYK
Sbjct: 237 HGRFVIVQVVDGIRRPKELIDQALAGKAFVY---GSGAQSGSDGAGGKPKLGFYK 292

Query: 298 HLMGGVSGMLPFVIGGGIMIAIPLFNIILGVKPDQISNLGVSHTIAALFNKIGGA-AFA 356
HLM GVS MLFPV+GGGI+IAI+F+P P D SYN A + IGG AF
Sbjct: 293 HLMGGVSNMLPFVVGSGGILIAISFMPIKAFDSDP-----SYHFAEMLMTIGGGNAG 347

Query: 357 FMLFVLGAYIAYSIAEKPGILVAGFVAGSISGLIAPGKVPFADGGKATIALAGVPSGFLG 416
M+FYLA +IA SIA++PG AG + G IAS+G A GFLG
Sbjct: 348 LMTFVLAAFIANSIARPGFANGMIGGLIATGEA-----GFLG 386

Query: 417 LAMVGGFLAGGVILLIRKLLSGPKLSLEGIKSILLYPLIGVLTGFMILVNIIPHAINT 476
L+ GFLAG V L +K+L+ LP+L+QIK+L YP+ + ITG +M++ P+AA NT
Sbjct: 387 GLDAGFLAGTVALGVKKVLANLPCTLDGIKTILFYFVFNPIFTGMIMLVIVGSLAFTNG 446

Query: 477 LMTFLQGLSGSSAVIMLLGVGMVAVDMGSPVHKAAVYVGTGTLAATVANSGSVMAAVM 536
L +L + ++ V+G++ GGMVAVDMGSP+KAA+ PG + A G AAVM
Sbjct: 447 LQKLGSMGTANMVILGVILGGMVAVDMGSPINKAA+PTGIAIDRA---GNFGPHAAVM 502

Query: 537 AGGVVPLPVLAVFATLLPKDKFNBERGSLQNTVMGLSPITEGALPPGAADPARAIPSI 596
AGGVVPLP + +AT LFK KF +ER++G TN +G SFITEGALPP AADP R IPS I
Sbjct: 503 AGGVVPLGLALATTLFKKFTVQREAGKTYWILGASPTTEGALPPAADAQRVIPSII 562

Query: 597 VGSALTGALVGLAGIKIMAPHGGIFVI---ALTSNPLLYFILGIAVGVGLPGLPKR 652
VGSAG I L L + L APHG FVI + +NPLLY++ I+ G++V+ +L G +K
Sbjct: 563 VGSAPAGGLTALFNVTLGAPHGAFVIFIGNIVNPLLYLVIAIAGSIVTALLGFWKK 621

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6053> which encodes the amino acid sequence <SEQ ID 6054>. Analysis of this protein sequence reveals the following:

Possible site: 18
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -10.77	Transmembrane	624 - 640 (612 - 646)
INTEGRAL	Likelihood = -7.59	Transmembrane	301 - 317 (297 - 321)
INTEGRAL	Likelihood = -6.85	Transmembrane	442 - 458 (439 - 468)
INTEGRAL	Likelihood = -5.95	Transmembrane	409 - 425 (406 - 426)
INTEGRAL	Likelihood = -3.61	Transmembrane	524 - 540 (523 - 547)
INTEGRAL	Likelihood = -2.50	Transmembrane	337 - 353 (337 - 353)
INTEGRAL	Likelihood = -2.44	Transmembrane	589 - 605 (589 - 605)
INTEGRAL	Likelihood = -1.70	Transmembrane	480 - 496 (480 - 500)
INTEGRAL	Likelihood = -1.44	Transmembrane	370 - 386 (370 - 386)

----- Final Results -----

bacterial membrane	---	Certainty=0.5310 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

The protein has homology with the following sequences in the databases:

>GP:BAB04547 GB:AP001510 PTS system, fructose-specific enzyme II, BC component [Bacillus halodurans]
 Identities = 322/659 (48%), Positives = 431/659 (64%), Gaps = 48/659 (7%)

Query: 1 MKIDLLRKDIMILDQAIKSEKVAIDEMITKLVEKDIVHDVDFVKGSIMTRESQTSGLG 60
+KI +LL+ED M+L+L+A SKE IDE+ L + +D FK++I RE Q++TG+G
Sbjct: 2 LKISELLKKITMVNLIRASKEAVIDELVRLTDKAGRINDAQAFKRAILLERESQSTTGWG 61

Query: 61 DGIAMPHSKNIVVDKPAVLPAKSNKGVGYDKALDQOPTDLFFMIAAPQGANITHLAALAE 120
+GIA+PH+K V +PA+ F +S+ G+DY++LDGQP+ LFFMIAA +GAN+ HD L+ L
Sbjct: 62 BGIAIPIAKTAAPVQKPAIFPGRSDAGIDYESLDGQPSPHLPFMIASBSGANNEHETLSRL 121

5 Query: 121 SQYLLKDGFPADKLRAAATPEAVIAVFD--EASTAKEEVVAPTSGQDFIVAVTACTPTGIAH 178
S+L+ + F L A + + +A D EA TA E + ++AVT CPTGIAH
Sbjct: 122 STPLMDETFRSLTAKQAQSEDELAIDDKKEAEATAGEAEEKQGYE--LLAVTGCPCTGIAH 179

10 Query: 179 TYMAEALKKQAASNGVAIKVETNGASGVANRLTASDIQRAKGVIAADKAVEMDRPCK 238
TYMA + LK +A E+G+IKVETNG+ GV NRIT B+I AK +IVAAD VEMDRF GK
Sbjct: 180 TYMAADNLKSKAQEIGVSIKVTNGSGGVKNRITREISAKAIIVAADIKVEMDRFNGK 239

15 Query: 239 QFIARFPVADGIKKSQELISLIANNBGNTYHAKNGKSETAVSTETKSLG-----AFYKHL 293
I PV DGI++ +ELI L A + Y + S E S GG FYKHL
Sbjct: 240 FVIQVPVTDGIRRPKLLIDQALAGKAPVY-----EGGAQASGEDGSAGGGRPKLGFYKHL 294

20 Query: 294 NGGVSQMLPFVIGGGIMIALFLDNLMLGVPNDQLGSLGSYHEIAAIFMNIIGA-AFSPM 352
N GVS MLFPV+GGGI+DA++P+ P+D SYH A + M IGG AF M
Sbjct: 295 MNGVSNMLPFVVGSGILLIATSPFGIKAPDPSDP-----SYHPAEMATTGGGAGFM 349

25 Query: 353 LPVLAGYIAYSIAKPKGLVAGFVAGAIASNGLAFKVPFAAGGEVSLGTGVPSPGLGAL 412
+PVL+ +IA SIA++PG AG + G IAS G A GFLG L
Sbjct: 350 IFLAAFIAMS IADRRFGAKMGIGLIASTGEA-----GFLGSL 388

30 Query: 413 VGGFIAGGVILALRKLLAGLPKSLGVKSSILLYPLLVGLVGTGFLMFPVNIPGAINTALN 472
+ GFLAG V L ++K+LA LP++L+G+K+IL YP+ + +TG +ML + P+AA NT L
Sbjct: 389 IAGFIAGYVALGVKKVLANLQCTLDGKTLTFLYFVNIPTTGMIMLVGSLPAAPNTLQ 448

35 Query: 473 DFLQSLGSSVALMGLLNGMMMAVDMGKPVNGAAVPGTGTAAATVANGSSVMAAVNAG 532
D+L + ++ V++G+++GMMMAVDMGKPVNGAA+ PG + A G AAVNG
Sbjct: 449 DWLGSMTANHVILGVLSGMMMAVDMGKPVNGAA+ PG + A G AAVNG 504

40 Query: 533 GMVPFLAVFATLLFKDKFTKEERESGLTNIWGLSPITEGAIPFGAADPARAIPSPFIAG 592
GMVPFL + +AT LFK KFTK+ERE+G TN ++G SPITEGALIP ADP R ITP I G
Sbjct: 505 GMVPFLGIALATLLFKKFTKQEREAGKNTIILGASFPITEGAIPFAAADGRVITSPFIAG 564

Query: 593 SALTGALVGLAGIKLMAAPHGSGFVI--ALTSNPILYLVFVIGALVSGILGALAKGA 648
SA G L L + L ADP33 FVI + +NP+LXLV ++ G+V+ +L G +K A
Sbjct: 565 SAPAGLTALFNVTLASPHGCAFPVIFIGNVNNPLLHLVLIAGSVTALLGKAPAKGA 623

An alignment of the GAS and GBS proteins is shown below.

Identities = 526/652 (80%), Positives = 581/652 (88%), Gaps = 6/652 (0%)

45 Query: 1 MKIQDLLKKKVMIMDLKATSKBAIAIDEMITKLVDVGVTNFAIPKDGIMCREAQSTGLG 60
MKIQDLL+K++ML+DL+A SKE AIDEMITKLV+ +V +F +PK IM RE QSTGLG
Sbjct: 1 MKIQDLLRKIDMILDLQAIKSKEVAIDEMITKLVEKIDVHDFVFKKSIMTREETSTGLG 60

50 Query: 61 DGIAMPHSKNAAVKEATVLPKASNGVDYALDQOPTDLFFMIAAPDGANITHLAALAE 120
DGIAMPHSKN V + VLPKAS NGVDY+ALDQOPTDLFFMIAAP GANITHLAALAE
Sbjct: 61 DGIAMPHSKNIVVDKPAVLPAKSNKGVGYDKALDQOPTDLFFMIAAPQGANITHLAALAE 120

55 Query: 121 SKYLLKDGFPADQLRQAATPDIIATFDSNSISQETVAPQTVQSTGSGSDYIVAVTACTIG 180
S+YLK+GFAD+LR A TP+ +IA FD S ++E V T G D+IVAVTAC TG
Sbjct: 121 SQYLLKDGFPADKLRAAATPEAVIAVFDASTAKEEVVAPT-----SQQDFIVAVTACTIG 175

60 Query: 181 IATYMAEALKKKQAASNGVAIKVETNGASGVANRLTSSDIQRAKGVIIAADKAVEMDRP 240
IATYMAEALKK+AAENG VIKVETNGASGV N+LT+ DI RAKGVI+AADKAVEMDRP
Sbjct: 176 IATYMAEALKKQAASNGVAIKVETNGASGVANRLTASDIQRAKGVIAADKAVEMDRP 235

65 Query: 241 DGRPLVSRFPVADGIKKSBDLNIILDNKAQTYHAKNNDKQSGRSIDKSLGSAFYKHL 300
DKK ++RPVADGIKKS++LI++IL+N+ TYHAKN ++ S K+ LG AFYKHL
Sbjct: 236 DGRQFIARFPVADGIKKSQELISLIANNBGNTYHAKN--GKSTAVSTETKSLGGAFYKHL 294

Query: 301 GGVSQMLPFVIGGGIMIALFLDNLMLGVPNDQLGSLGSYHEIAAIFMNIIGA-AFSPM 360
GGVSQMLPFVIGGGIMIALFLDNLMLGVPNDQLGSLGSYHEIAA+F NGGAAP+PFLP
Sbjct: 295 GGVSQMLPFVIGGGIMIALFLDNLMLGVPNDQLGSLGSYHEIAAIFMNIIGA-AFSPM 354

-2199-

Query: 361 VLAGYIAYSIAEKPGLVAGFVAGSIASSGLARIVKPPFAEGGKATLALAGVPSGFLQALVG 420
 VLAGYIAYSIAEKPGLVAGFVAG+IAS+GLAIGKVPFA QG+ +L L GVPSGFLQALVG
 Sbjct: 355 VLAGYIAYSIAEKPGLVAGFVAGAIASGLAIGKVPFAAGGSEVSLGTVPSGFLQALVG 414

5 Query: 421 GFLAGGVILLRLKLLGSLPRSLBGIKSILLYPLLVITGFLMILVNIIPMAINTALNIF 480
 GFLAGGVIL LRKLL+GLP+SLG+KSILLYPLLVITGFLMIL VNIIPMAINTALN F
 Sbjct: 415 GFLAGGVILALRKLGLPRSLBGVKSILLYPLLVITGFLMILFVNIIPMAINTALNDF 474

10 Query: 481 LQGLSGSSAVLMGILLVGGMVAVDMMGGFVNKAAYVFGTGTLLAANTVANGGSSVMAAVNAGGM 540
 LQGLSGSSAVLMGILLVGGMVAVDMMGGFVNKAAYVFGTGTLLAANTVANGGSSVMAAVNAGGM
 Sbjct: 475 LQGLSGSSAVLMGILLVGGMVAVDMMGGFVNKAAYVFGTGTLLAANTVANGGSSVMAAVNAGGM 534

15 Query: 541 VPFLAVFVATLLFKDKFKNBERQSGLTINIVMGLSPFITGCAIPFGAADPARAIPSFIVGSA 600
 VPFLAVFVATLLFKDKF EER+SGLTINIVMGLSPFITGCAIPFGAADPARAIPSF I GSA
 Sbjct: 535 VPFLAVFVATLLFKDKFTEERESGLTINIVMGLSPFITGCAIPFGAADPARAIPFVLAGSA 594

20 Query: 601 LTGALVGLAGIKLMAFHGGIFVIALTSNPLLYILFILGAVVSGVLEGLFRK 652
 LTGALVGLAGIKLMAFHGGIFVIALTSNP+LY++F++IGA+VSG+LFG RK
 Sbjct: 595 LTGALVGLAGIKLMAFHGGIFVIALTSNPLLYLVFVIGALVSGILPQALG 646

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1954

A DNA sequence (GBSx2063) was identified in *S. agalactiae* <SEQ ID 6055> which encodes the amino acid sequence <SEQ ID 6056>. Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.1532 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:AAC24914 GB:AF012285 fructose-1-phosphate kinase [Bacillus subtilis]
 Identities = 146/303 (48%), Positives = 197/303 (64%)

Query: 1 MIYTVTLNPSIDFIVRLDTLLGSGVMRMTSDDKYVGGKGINVSRILKRLKIDNTATGFIG 60
 MIYTVTLNPS+D+IV ++ +G +NR + D KY GGGKGINVR+LNR + + A GF+G
 40 Sbjct: 1 MIYTVTLNPSVDYIVFVEDF+VGLNRSSYDTKYPGGKGINVSRILRHRHVASKALGFVG 60

Query: 61 GFTGHFVEDGLVLBGIKTDPSVSNEDTRINKVKAKITEINGGGFRITNEQLHRLEKLL 120
 GFTG +++ L E ++T F V DTRINK+K ETEING GP I++E +
 45 Sbjct: 61 GFTGEYIKTFLRSENLETAFSVKGEDTKINVKLTGDETEINGGPTTSDSDPKALRQF 120

Query: 121 SRLTFEDTVVFGAGSAPASLGNKYNVTLPIAKTKGAEVCDPFGQTLIDALAYQPLLVKP 180
 L D VV AGS P+SL + Y + K+ A VV D G+ LL A +P L+KP
 50 Sbjct: 121 QSLQSGSDIVLWAGSIPSSLPDHTYRKIAEACQCNARVVDLISGAILLKATRMKPFLLMKP 180

Query: 161 NNHELADIFGVELBGLPDISKYAHKILDKAGKKNVISMGGDGLLVTPESAYFAPKIGKE 240
 N+REL ++FG + + + Y K+++GA++VIVSMGGDGLL T RA YFA KG+
 55 Sbjct: 161 NNHELGMFGTALTSVEEAAPYKGLVQGAHVIVSMGGDGLLFTNEAVYFANVFGK 240

Query: 241 VKMSVNGSDSMVAGFTGKFPKVNPRALKWVACGTATTFSSDGLATAPFQDIYNKVEV 300
 + NSVNGSDS+VAGF K EA + GV G+AT FS++L T EF+Q + +V+V
 60 Sbjct: 241 LVNSVNGSDSVVAGFLAGISKQLPLRAFLGLVTSGSATAPFSBELGTBEFVQQLPEVKV 300

Query: 301 EKL 303
 +L
 60 Sbjct: 301 TRL 303

-2200-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6057> which encodes the amino acid sequence <SEQ ID 6058>. Analysis of this protein sequence reveals the following:

```

Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1738 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 222/302 (73%), Positives = 261/302 (85%)

Query: 1 MIYTVTLNPSIDFIVRLDTLLGSGVNMNTSDKKYVSGKGINVSRILKRLKINDATGTFPG 60
MIYTVTLNPSIDFIVR+D+ LGSVMNM SDDK+ GGGKGINVSRIL+RL I +TATGF+G
Sbjct: 1 MIYTVTLNPSIDFIVRIDQINLGSVNMMSDDKFAGGKGINVSRILQRLDIATATGTFGLG 60

Query: 61 GFTGHFVEDGLHLEGIKIDFVSVNEDTRINVVKAKIETENGSGSPRITNEQLHLEKLL 120
GFTG F+E+ L EG+KIDFV ++DTRINVK+K++ ETE+NG GP I+ EQL L+ L
Sbjct: 61 GFTGRFIEREGLSAGSVKTDIPVKGQDQTRINVVKIKSQETELNQGPITISGEQLDELTKL 120

Query: 121 SRITPREDTVVFAGGAPASLGKVKYNTLPIAKKTGAEVVCDPFGQTLIDALAYQPLLVKP 180
S+LT EDTVVFAGGAP+LGN VY L+P+ ++GA+VVCDFEGQTL+DALAY FLVVKP
Sbjct: 121 SQTAREDTVFAGGAPANLGNVYKLLPLVRQSGAQVVCDFEGQTLIDALAYNPLLVKP 180

Query: 181 NNHELADIPGVEGLPDIKVAHKILDKGAKNVVSMAGDGALLVTPESYFAKPIKGE 240
NNHEL IFG L L D+E YA ++L+ GA-NVI+SMAGDGALLVTP EA-YFAKPIKGE
Sbjct: 181 NNHELEAIPGTILTSILDVETIYARRILLEMSAQNVII+SMAGDGALLVTPKATYFAKPIKGE 240

Query: 241 VKNSVVGAGDSMVAGFTGEFVSKKIPVEALKWGVACGTATTFSDILATASFICQYINKVSV 300
VKNSVVGAGDSMVAGFTGEF+KS+NP+EALKWGVACGTAT FSDILAT FI++ Y+KVEV
Sbjct: 241 VKNSVVGAGDSMVAGFTGEFVFMSQNPFEALKWGVACGTATFSDILATIAFIKETYHKVSV 300

Query: 301 EK 302
EK
Sbjct: 301 EK 302

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 1955

A DNA sequence (GBSx2064) was identified in *S.agalactiae* <SEQ ID 6059> which encodes the amino acid sequence <SEQ ID 6060>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2769 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9961> which encodes amino acid sequence <SEQ ID 9962> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GF:AAC24913 GB:AF012285 FruR [Bacillus subtilis]
Identities = 97/247 (39%), Positives = 148/247 (59%), Gaps = 4/247 (1%)

Query: 23 MLKSKRKEITLSRLSEQNKSVTLDELTSILFSTSESTVRRDLDELESAGFLKRVHGGAEIPY 82

```

-2201-

- ML +R ++I+ ++E++ V + EL ++ SEST+RRDL LE OFLKRVBGGA
 Sbjct: 1 MLYPERHQI I I I Q I E K H D V V K I Q E L I N I T N A S E S T I R R D L S T L E E R G P L K R V H G G A A K L S 60
- Query: 83 SLQQLSNQSEFAIKNVQKLDIARQTAFLIAKQDVIFIDAGTTTELLIDFLPH-BQJLVV 141
 + E EK+ KH+ KL IA + A L+ + D I++DAGITT +IDF+ + + VV
 Sbjct: 61 DIRLEPMLERKSSNKLHDKLKIARQAASLLEEGDCIYLDAGTTTLHMIDFMDKTDIVV 120
- Query: 142 TNSIHAHAKLVDRGIKTTIIGGAVKHSIDASIGQVAINQIRQITVDKAPLGMNGID-KVY 200
 TN + H L+ + I ++GG VKH T A IG ++ + Q DK+FLG NG+ E
 Sbjct: 121 TNGVMHIDALIRKISFLYLGGYVKIRTGALIGASLAVMDQYRFKRSFLGTNGVHTTAG 180
- Query: 201 LTTTDLSEAAIKEAIINNSQQTFLMDSKIGQVTFKRVKREINDINLVNKTDSHLMVII 260
 TTTF +EA +K+ I ++ ++L D SK G+++F+ I D ++T TD+E +T
 Sbjct: 181 FTTTDPDEALLKQAKINQAKHAYVLADPSKFGESISFSAFAGIGDATTIIT--TDMEKLTTF 238
- Query: 261 KEKMKVI 267
 + K +
 Sbjct: 239 NYQKTV 245
- 20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6061> which encodes the amino acid sequence <SEQ ID 6062>. Analysis of this protein sequence reveals the following:
- Possible site: 27
 >>> Seems to have no N-terminal signal sequence
- 25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2604(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
- 30 An alignment of the GAS and GBS proteins is shown below.
- Identities = 135/237 (56%), Positives = 184/237 (76%)
- Query: 33 LSRLEQNKSVTLDELTSILETSBSTVRRDLDELESAGFLKRVHGGAEPLFSLQQLSNQSE 92
 ++++ + V+L+L +L +SEST+RRDL ELE G L RVHGGAE L +SL +ELSNQE
 Sbjct: 1 MAKITEENTVSLIEDMLQLNNSSESTIRRDIGLEQEGRLHVRVHGGAEPLFSLQQLSNQSE 60
- Query: 93 KAIRNVQKLDIARQTAFLIAKQDVIFIDAGTTTELLIDFLPHEQLVTVNTHHAARLV 152
 K++N K IA++ ++LI DVIFIDAGTTIE L+ FL + LTVNTHHAAR+LV
 Sbjct: 61 KSVNTHSHIKKAIAQASQLIYNDVIFIDAGTTTEFLPLPQAQNLTVTVNTHHAARLV 120
- Query: 153 DRGIKTTIIGGAVKHSIDASIGQVAINQIRQITVDKAPLGMNGIDVYLLTTTDLSEAAIK 212
 + I+TII+GE VK +TDASIG VA+ QIRQ+ DKAFLGMNG+E+ YLTTFD+EA IK
 Sbjct: 121 ELSIETIIVGGYKQVTTDASIGVALEQIRQNFQDKAFLGMNGVDYSYLTTFDMEAEVVK 180
- Query: 213 EAINNSQQTFILMDSKIGQVTFKRVKREINDINLVNKTDSHLMVIIKEKMKVI 269
 +A++N++ +IL+D +KIQV+E KV IND+ ++T + ++ IKER KVI++
 Sbjct: 181 KAVLENAKIAYILVDGDKIQVSPVKVPAPNDVTIITLGGASILKQIKERKAVIEL 237

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1956

A DNA sequence (GBSx2065) was identified in *S.agalactiae* <SEQ ID 6063> which encodes the amino acid sequence <SEQ ID 6064>. This protein is predicted to be beta-lactam resistance factor. Analysis of this protein sequence reveals the following:

- Possible site: 32
 >>> Seems to have no N-terminal signal sequence
- Final Results -----
 bacterial cytoplasm --- Certainty=0.5777(Affirmative) < succ>

-2202-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 5 >GP:CA89121 GH:AJ277485 beta-lactam resistance factor
(*Streptococcus pneumoniae*)
Identities = 215/410 (52%), Positives = 283/410 (68%)
- 10 Query: 1 MTLRELTIIEFEKHSQNYDSQSFLQTPBMAKLEKRGYDVRLGYQVFNKLEISLSYIM 60
M L LT EEF+ +S S+SF+Q+ +M LLEKRG + YL + E ++++ +L Y +
Sbjct: 1 MALTTLTKEEFQTSQVSSRSFQSVQMSDLEKRGARIVYIALKQEGRIQVAALVYSL 60
- Query: 61 PVIIGFQMKIDGGPVHSNSKYLKQFYKALQYAKNSGVLELIVEFYDYQLPTSSGVPSN 120
P+ GG M+++SGP+++ L FY L+ YAK NGVLEL+V+FY+ YQ F S G P +
15 Sbjct: 61 PNIIGLHNEINSGPIYQQDALPVFYFARLKRYAKQNGVLELLVKPFTYTYQTFDSQGNPID 120
- Query: 121 QGNQNLIEDFTSSGYHHDGLTGTGFKYLSMRYKNLEGVITSETLLSSFSKTRGALVKA 180
++I+D T GY DGLTGT+ G W Y K+L +T ++LL SPK G+ LVKKA
Sbjct: 121 AEKSKIIQDLTDLGYCFDGLTGTGYPGEPMLYYKDLTELTEKSLKSKKGPVLVKA 180
- Query: 181 MSPGIKVRVLKRDELHFLKEITTSNKRNDYNDKSLDYQDFYDSFEGKAEFVIATINFR 240
+FGI+++ LKR+EL +FK IT TS RR+Y DKSL+Y+ FYD+F +AEF+IA+LNP
Sbjct: 181 ETFGIRLKKLKEELSIKFNITKETSRRREYSDKSLLEYEHFYDTPCEQAEPLIASLNF 240
- 25 Query: 241 EYDHNQIKABALENKLKLLDERFRENADSPKYHRQSEIINQLASFTRKQEVQSFQK 300
+Y LQ + LE L L +N S K Q E +Q +FE R+ E + I+K
Sbjct: 241 DYMKGQGEQSKLEENLDKLLDLSKNPSEKKQNLQREYSSQFETFEVRKABARDLIEK 300
- Query: 301 YDNQDVVLAGSLFVYSKETYVFFSGSYTEPNKFYAPVLQYVMQEALEKRGSTFFYNLLG 360
Y +D+VLAGSLFVY +ET Y PSGSYTEPNKFYAPA+LQ+YVM E++KRG YN LG
30 Sbjct: 301 YGEEDIVLAGSLFVYMPQETTYLPSGSYTEPNKFYAPALLQKYVMLESIKRGIPKYNFLG 360
- Query: 361 IQGTFDGSDSILRPFQNFNGCIIIRKMGTFNPNYSPFKYGIQLLKKVLKR 410
IQG FDGSD +LRPFQNFNG I+RK GTF Y+PSP KYK IQLLKK++ R
35 Sbjct: 361 IQGIFDGSGLVRPFQNFNGYIVRKAGTFRYHPSPLKYKAIQLLKKIVGR 410

There is also homology to SEQ ID 5460.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 1957

A DNA sequence (GBSx2066) was identified in *S.agalactiae* <SEQ ID 6065> which encodes the amino acid sequence <SEQ ID 6066>. This protein is predicted to be cell wall protein, 40 kDa (sr 5' region).

Analysis of this protein sequence reveals the following:

- 45 Possible site: 42
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -3.45 Transmembrane 25 - 41 (23 - 42)
- Final Results -----
bacterial membrane --- Certainty=0.2381 (Affirmative) < succ>
50 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9963> which encodes amino acid sequence <SEQ ID 9964> was also identified.

55 The protein has homology with the following sequences in the GENPEPT database.

IGB:AF278686 choline binding protein D; CbpD [Strept...]
IGB:AF278686 choline binding protein D; CbpD [Strept...]

-2203-

>GP:AAF87768 GB:AF278686 choline binding protein D; CbpD
[Streptococcus pneumoniae]
Identities = 63/230 (27%), Positives = 108/230 (46%), Gaps = 34/230 (14%)

5 Query: 324 WTEQGGQDDIKWYTAVTGSG-----NYKVAVSFADHKNEKGLYNHLYYQASGTLVG 377
W+ G + W + V GD NY S+ +++++ G VG
Sbjct: 123 NSTAGTYGHVAVVSNVM-GDQIEIEEYNYGTTSEYKVRKIKANTMTGPIHFKDLDGGSVG 181

10 Query: 378 VTGTVTVAGTNSSQEPIENGAKTGVYNIIGSTEVIKNEAKISSQTPTLEKGDKNINYDQ 437
+ + + GT+ + + +K E S G+K++YDQ
Sbjct: 182 NSQGSTSTGGTHYFKT-----KSAIKTEPLASGTVIDYYPGKRVHYDQ 225

15 Query: 438 VLTADGYQWISYKSYSGVRRYIPVKKLTTSEKAKDEATKPTSYNLPKGT-TYTFKTV 496
+L DGY+W+SY +Y+G RY+ ++ + + P L TG T+ F
Sbjct: 226 ILEKDGKWLSTYATNGSYRYVQLEAVNKN-----PLGNSVLSTGGTHYFKTKS 275

20 Query: 497 DVKSQPKVSPFVERNPQKGEKIHYDQVLVVDGHWISYKSYSGIRYIEI 546
+K++P VS+ V + GK+HYDQ+L DG++W+SY +Y+G RRYI++
Sbjct: 276 AIKTEPLVSATVIDYYPGKRVHYDQILEKDGKWLSTYATNGSRRYIQL 325
Identities = 49/161 (30%), Positives = 85/161 (52%), Gaps = 14/161 (8%)

25 Query: 116 GNTVYSKETEVIGWTFPSKAPVAFYAKNGDKVFDQVFNKDNWIKWISYKSPGVRRYAAIE 175
G + + ++ K P S V Y G+KV YDQ+ KD K+W+SY ++ G RY +E
Sbjct: 191 GTHYFKTKSAIKTEPLASGTVIDYYPGKRVHYDQILEKDGKWLSTYATNGSRRYVGL 250

30 Query: 176 SLDPGSGSETKAPTPTVINSQSNQEKIATQGNFTPSHKVEVQKHAASPTQFTLDKEDR 235
+++ + + + P+ RS + + + +T G+ F K +K E V++ G++
Sbjct: 251 AVNKN-----PLGNSVLSTGGTHYFKTKSAIKTEPLVSATVIDYYPGKE 296

35 Query: 236 IFYDQILITBGNQWLSYKSPFNQVRFULLKASVSEKTEDK 276
+ YDQIL +G +WLSY ++NG RR++ L +S + +++
Sbjct: 297 VHYDQILEKDGKWLSTYATNGSRRYIQLBQVTSQGVQWQ 337
Identities = 52/192 (27%), Positives = 90/192 (46%), Gaps = 13/192 (6%)

40 Query: 295 ISNETTTGFDILITNIDKNGIAAVKVPVWTEGGQDDIKWYTAVTGSGNYKVAVSFAD 354
I T TGF + KD +G + T GG K +A+ T + +
Sbjct: 161 IKANTMTGPF----IHFKDLDGGSVNSQGSTSTGGTHYFKTKSAIKTEPLASGTVIDY 215

45 Query: 355 HNEKGLYNHLY---YQASGTLVGVTKTVVAGTNSSQEPIENGAKT--GVYNIIG 409
+ EK Y+ L Y+ S T + V+ N+ P+ N+ + G+
Sbjct: 216 YPGKRVHYDQILEKDGKWLSTYATNGSYRYVQLEAVNKN--PLGNSVLSTGGTHYFKT 273

50 Query: 410 STEVIKNEAKISSQTPTLEKGDKNINYDQVLTADGYQWISYKSYSGVRRYIPVKKLTTSE 469
+ +K E +S+ G+K++YDQ+L DGY+W+SY +Y+G RRYI ++ + TSS+
Sbjct: 274 KSAIKTEPLVSATVIDYYPGKRVHYDQILEKDGKWLSTYATNGSRRYIQLBQV-TSSQ 332

55 Query: 470 KAKDEATKPTSY 481
++++ +SY
Sbjct: 333 NYQNSGINSY 344
Identities = 33/113 (29%), Positives = 56/113 (49%), Gaps = 2/113 (1%)

60 Query: 91 NTATKDIITPLVETKPMVEKILPEQGNVYYSK-ETEVKNTPSKAPVAFYAKNGDKVFDY 149
N + + + V P+ L G Y K + +K P SA V Y G+KV YD
Sbjct: 241 NGSYRVQLEAVNKNPLGNSVLSTGGTHYFKTKSAIKTEPLVSATVIDYYPGKRVHYD 300

Query: 150 QVFNKDNWIKWISYKSPGVRRYAAIESLDPGSGSETKAPTPTVINSQSNQEKI 202
Q+ KD K+W+SY ++ G RRY +E + S + + + ++ GS++ +
Sbjct: 301 QILEKDGKWLSTYATNGSRRYIQLBQVTSQGNVYYSKINSYSGHSSSTV 352

A related GBS gene <SEQ ID 8937> and protein <SEQ ID 8938> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
McG: Discrim Score: -6.74
GVH: Signal Score (-7.5): 1.26
Possible site: 42

-2204-

```
>>> Seems to have no N-terminal signal sequence
ALOM program count: 1 value: -3.45 threshold: 0.0
  INTEGRAL Likelihood = -3.45 Transmembrane 22 - 39 ( 23 - 42)
  PERIPHERAL Likelihood = 6.26 371
modified ALOM score: 1.19
```

*** Reasoning Step: 3

```

----- Final Results -----
      bacterial membrane --- Certainty=0.2381(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear)

```

The protein has homology with the following sequences in the databases:

41.2/57.9% over 283aa

cus mutants

EGAD 33594	cell wall protein, 40 kDa (sr 5' region)	Insert characterized
PIR A60328 A60328	40K cell wall protein precursor (sr 5' region) - (strain OMZ 175, serotype f)	Insert characterized

ORF02145(301 - 1803 of 2238)
EGAD|33594|34911(30 - 313 of 335) cell wall protein, 40 kDa (sr 5' region) [Streptococcus mutans]PTR|A60328|A60328 40K cell wall protein precursor (sr 5' region)
)- Streptococcus mutans (strain OMZ175, serotype f)
%Watch = 8.0
%Identity = 41.1 %Similarity = 57.9
Matches = 81 Mismatches = 79 Conservative Sub.s = 33

156 186 216 246 276 306 336 366
 *YA***FCTKNKSWVFFSRISYSIKYYICITNISIKC*HVTKRIL**CK*IRK*VFNMKGQVNDTKQSYSLRKK
 : : : : : : : :
 MNQKIVISSFYMLGAHSPSKAVYHNRSVKIMKRIDINHQAQRFSIRKYA
 10 20 30 40 50

[illegible]

654 684 714
ETKPMVEK-----TLPEQNYVYSKETEVNTPSKAPVA
| : | ::|||::| |||
EVEEKSEKVNPALPNVDKGAKSKEDVN----AEQNEKAVERNLMCRQAQAVSPGQNYVFOETTPVNAAMSSP-
130 140 200 210 230 240

[illegible]

1713 1743 1773 1803 1833 1863 1893 1923
 KGLTTSSEKADKPTKTSYINLPKIGYITFTTKTVDSQPKVSSPVEFNPKQGEKIHYYQVLVDGHHMISYKYSIGIR
 :
 -----AVTIEELKKKEIVQONLPAGQTHYFTKQOOLMKMKLNCILVRNSPRTFTTITFFMIRF
 290 300 310 320 330

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6067> which encodes the amino acid sequence <SEQ ID 6068>. Analysis of this protein sequence reveals the following:

Possible site: 23

```
>>> Seems to have a cleavable N-term signal seq.
```

-2205-

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5

The protein has homology with the following sequences in the databases:

>GP:AA67768 GB:AF278686 choline binding protein D; CbpD
 [Streptococcus pneumoniae]
 Identities = 93/217 (42%), Positives = 136/217 (61%), Gaps = 18/217 (8%)

Query: 42 GDNTPSKWKKGNG-IDSNWMIROCTSFAPFLSSANGPOLPKGYGNACTWGHIAKNQGY 100
 GD+YP+ +K G+ ID W MY RQCTSF AFRLS+ NGF++P YGNA WGH A+ +GY
 Sbjct: 51 GDDYPAYYKNGSQEIDQNRMSRQCTSFVAFRLSNVNGFELPAAYGNANWGHIRARRBGY 110

Query: 101 PVNKTPTSIGAIANWFDKNAYGSNAAYGHVAVVADIRGDTVTIEBZYNNAGQSPRYHKRQI 160
 V+ TP+IG+I W + YGHVAVV++ GD + IEEYNY E Y+KR I
 Sbjct: 111 RVDNTPITGSITW-----STAGTYGHVAVVSNVMDQIEIEBZYNYG---TESYMKRVI 161

Query: 161 PKSQVSGYIHFKDLSSQTSHSYPRQLKHISQASFDPSGTYHFTTTLPLVKGQTSIDSPDLA 220
 + ++G+IHFKDL + + SQ+S GT++F T+ +K + +
 Sbjct: 162 KANITMTGFIHFKDLGGSGVN-----SQSSTGTGTHYFKTSIAIKTEPLASGTVID 213

Query: 221 YYEAGQSVYYDKVVTAGGYTWLSYLSFGSNRRYIPIK 257
 YY G+ V+YD++ GY WLSY ++G+ RY+ ++
 Sbjct: 214 YYPGKGVHYDQILEKDGYSKWLSTAYNGSYRYVQLE 250

10

15

20

25

An alignment of the GAS and GBS proteins is shown below.

Identities = 34/94 (36%), Positives = 52/94 (55%)

Query: 453 SGVRRYIPVKKLTTSEKAIDEATKPTSYPNLPKTKGTYTFTKTVGVKSGPKVSGPVEFNF 512
 S V YI K L++ + + K S + +GTY FT + VK Q + SP +
 Sbjct: 163 SOVSGYIHFKDLSSQTSHSYPRQLKHISQASFDPSGTYHFTTTLPLVKGQTSIDSPDLAY 222

Query: 513 QKGEKIHVDQVLVVDGHWISYKSYSGIRRYIEI 546
 + G+ ++YD+V+ G+ W+SY S+SG RRYI I
 Sbjct: 223 EAGQSVYYDKVVTAGGYTWLSYLSFGSNRRYIPI 256
 Identities = 30/78 (38%), Positives = 45/78 (57%), Gaps = 2/78 (2%)

Query: 402 TGVNIIIGSTEVIKNEAKISSQTOFTLEKGDKINYDOVLTAAGQWISYKSYSGVRRYIPV 461
 +G Y+ VK + I S E G + YD+V+TA GY W+SY S+SG RRYIP+
 Sbjct: 197 SGTYHFTTTLPLVKGQTSIDSPDLAYEAGQSVYYDKVVTAGGYTWLSYLSFGSNRRYIPI 256

Query: 462 KKLTTSEKAIDEATKPT 479
 K+ + +++ TKP+
 Sbjct: 257 KE--PAQSVVQNDNTKPS 272
 Identities = 27/94 (28%), Positives = 47/94 (49%)

Query: 198 NQEKIATQGNVYFSHKVBEVKNKVAESPTQFTLDKGRIDFQDILIEGNQWLSYKSPNG 257
 +Q G Y F+ ++ VK + + SP + G ++YD+++T G WLSY SF+G
 Sbjct: 190 SQASFDPSGTYHFTTTLPLVKGQTSIDSPDLAYEAGQSVYYDKVVTAGGYTWLSYLSFG 249

Query: 258 VRRFVLLGKASSVEKTEDEKESVPOPARITKIG 291
 RK++ + + + D K S + +T G
 Sbjct: 250 NRRYIPIKEPAQSVVQNDNTKPSIKVGVITVTFPG 283
 Identities = 23/73 (31%), Positives = 35/73 (47%)

Query: 103 ETGKMWERTLPEQGNVYVSKETEVKNTKPSAPVAFYAKKGDKVFDVFNKDNVKNISY 162
 + K + + + G Y ++ VK S +P Y + G V+YD+V W+SY
 Sbjct: 185 QLKHISQASFDPSGTYHFTTTLPLVKGQTSIDSPDLAYEAGQSVYYDKVVTAGGYTWLSY 244

Query: 163 KSPGCVRRYAAIR 175
 SF G RRY I+
 Sbjct: 245 LSPSGNRRYIPIK 257

60

-2206-

SEQ ID 8938 (GBS91) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 18 (lane 7; MW 63kDa).

The GBS91-His fusion product was purified (Figure 195, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 283), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1958

A DNA sequence (GBSx2067) was identified in *S. agalactiae* <SEQ ID 6069> which encodes the amino acid sequence <SEQ ID 6070>. This protein is predicted to be thiamine biosynthesis protein. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0984 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB49673 GB:A0748285 PROBABLE 2-DEHYDRO-PANTOATE 2-REDUCTASE (EC
1.1.1.169) [Pyrococcus abyssi]
Identities = 85/301 (28%), Positives = 150/301 (49%), Gaps = 7/301 (2%)

Query: 1 MIVYIAGSGAMGCRPGYQISKTHDVILLNADHIMAIKENGLKVTGPTDILV/KLPIMK 60
M +YI G+QA+G FG ++ DV+L+ H+ AI E GLK+ G + VK+
Sbjct: 1 MKIYILGAGAI GSLFGILLANAGEDVLLIGR-DPHVSAINEGLKIIVGIKDLNVKVEATT 59

Query: 61 PTDATTEADLIILPTKAMQLPNMLQDIKKIKGKTNVLCILNGLGHEDVIRQYIPEHNIL 120
E+ DLI+L TK+ L+ + I+ K + VL + NG+G+ED I ++ +
Sbjct: 60 RVPE-EKPDILVLTAKSYSTIEALKSARHIV-KGSWVLSIQNGIEDKIIIEF--GGKAI 115

Query: 121 MGVTVTWATGLKPGHAGLEGVGSVNLQSIDPNQAGHRVTELLNEAKLQATYDENVLPN 180
G+T A ++ FG G G + ++ +V ++ N A ++ EN++
Sbjct: 116 GGITTINGAMVEAPGVIKWTKGVTIIGLYPQSKKEFKVADVFNSADIETHVSENIISW 175

Query: 181 IWRKACVMTGNTSTCALLDCTIGQLFASSEDGVNMVRIIEHFVTVGKAGVLEDEEITK 240
IW KA VN +N LL+ + ++ +M E++ E V G+E D +
Sbjct: 176 IWAKAIVNSAINPIGTILLEVRKVKVIRENDFLLNAGVVKEGCRVALQNGIEFDVPPMDL 235

Query: 241 VMDKTSVKAHHYPSMHQDLVQNRILTIDFLNCAVNRKGNLGDITPYRLTLPIHTKE 301
+ T + +Y SM OD+ ++ TE+D+NG + + + + + P L+ LI KE
Sbjct: 236 P-FQTLQETREYNYSMLQDIWRGKK-TEVDYINGKIVEYAKAVNLEAPNILLGLIKKE 294

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6071> which encodes the amino acid sequence <SEQ ID 6072>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1392 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-2207-

Identities = 262/307 (85%), Positives = 288/307 (93%)

Query: 1 MLVYIAGSGAMGCKRFGYQISKTHDVIILLDNWADHIMAIKENGKLVPTGDETELKVLPIMK 60
 5 Sbjct: 1 MLVYIAGSGAMGCKRFGYQISKTHDVIILLDNWADHIMAIKENGKLVPTGDETELKVLPIMK 60

Query: 61 PTDATREADLILPTKAMQLPNMLQDIKKIIGKTKVLCCLLGLGHEVDVIRQYIPEHNIL 120
 10 Sbjct: 61 PTDATREADLILPTKAMQLPNMLQDIKKIIGKTKVLCCLLGLGHEVDVIRQYIPEHNIL 120

Query: 121 MGVTVWVAGLGKSPGHAELGKGVSMQLQSIDPNNQKAGHRVTELLNEAKIQATYDKNVLPN 180
 15 Sbjct: 121 MGVTVWVAGLGKSPGHAELGKGVSMQLQSIDPNNQKAGHRVTELLNEAKIQATYDKNVLPN 180

Query: 181 IWRKACVNGTMMSTCALLDCTIGQLFASSEDGVMMVHEIIEHFVTVGKAGSVELDEREITK 240
 20 Sbjct: 181 IWRKACVNGTMMSTCALLDCTIGQLFASSEDGVMMVHEIIEHFVTVGKAGSVELDEREITK 240

Query: 241 YVMDTSVKAARHYPSMQRDLVQNRLLTEIDFLNGAVNKGEMLGIDTYPGRLLITQLIHTK 300
 25 Sbjct: 241 YVMDTSVKAARHYPSMQRDLVQNRLLTEIDFLNGAVNKGEMLGIDTYPGRLLITQLIHTK 300

Query: 301 ENVLISIK 307
 Sbjct: 301 EAVLINQ 307

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1959

30 A DNA sequence (GBSx2068) was identified in *Sagalactiae* <SEQ ID 6073> which encodes the amino acid sequence <SEQ ID 6074>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

35	INTEGRAL	Likelihood = -3.03	Transmembrane	61 - 77 (61 - 78)
	INTEGRAL	Likelihood = -1.33	Transmembrane	80 - 96 (79 - 96)

----- Final Results -----

	bacterial membrane	--- Certainty=0.2211(Affirmative) < succ>
	bacterial outside	--- Certainty=0.0000(Not Clear) < succ>
40	bacterial cytoplasm	--- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1960

A DNA sequence (GBSx2069) was identified in *S.agalactiae* <SEQ ID 6075> which encodes the amino acid sequence <SEQ ID 6076>. This protein is predicted to be regulatory protein (pfoS/R). Analysis of this protein sequence reveals the following:

50 Possible site: 49
 >>> Seems to have no N-terminal signal sequence

	INTEGRAL	Likelihood = -9.82	Transmembrane	317 - 333 (304 - 335)
	INTEGRAL	Likelihood = -7.64	Transmembrane	187 - 203 (183 - 217)
	INTEGRAL	Likelihood = -5.26	Transmembrane	24 - 40 (18 - 44)
55	INTEGRAL	Likelihood = -5.04	Transmembrane	143 - 159 (139 - 161)

-2208-

```

INTEGRAL    Likelihood = -2.34    Transmembrane  116 - 132 ( 115 - 136)
INTEGRAL    Likelihood = -2.13    Transmembrane   55 - 71 ( 55 - 71)
INTEGRAL    Likelihood = -0.96    Transmembrane  268 - 284 ( 268 - 284)

```

```

5  ----- Final Results -----
      bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

10 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AC65034 GB:AE001189 regulatory protein (pfoS/R) [Treponema
pallidum]
Identities = 138/358 (38%), Positives = 220/358 (60%), Gaps = 18/358 (5%)

```

```

15 Query: 2   T N T V T P K S T A G S F I N K V L G S T A T A I V V A L I P N A I L A T F L K P F L S Y G - L A A E F L H I V Q V F Q 60
      T ++P++   F + K + L G ++ IV+ L+P A I   +   L A H+V Q
Sbjct: 3   T Q S L S P R Q ---- F M M K I L N S S A G I V I G L V P P A I A G E L F R A L A P L S P L F A A L Y H V V L E T Q 58

```

```

20 Query: 61   F F T P I A G F L I G Q Q K F T P M Q Q L A V G A A Y I G S G M A Y T E V I Q K G V A T G S P O L R G I G D L I 120
      F P + G L+G Q F + + + + I S G + G++ + G I G D+I
Sbjct: 59   F S V P A L I G T I V L G L Q F H C S A F E V A T L A F V S V I A S G ----- N V T L Q N G A M L I T G I G D V I 110

```

```

20 Query: 121  N M M L A L A V L A V K V F G N K F G S L T I I L L P I I I G T G V G L G W K L L P V S Y V T T L I G Q G I N S 180
      N+ML +ALA++ V+ K G S L T I I L P+I+ G +G L P Y V +T +G+ I +
25 Sbjct: 111  N V M L S A L A I I L V R A L R K G L S L T I A L P V I V A V A G G V G S F S L P Y V K M I T L F V G R V I A T 170

```

```

25 Query: 181  F T T L Q P I A M S I L I A M A F S M L I V S P I S T V A I G L A I G L N G S A S A S M S V A S T T A V L W A T M 240
      F L Q P+ M S I L++M+ F S++I+ S P+S+V A+G+A+G L G++ A A++G V+S L+ T M
25 Sbjct: 171  F I A L Q P L I M S I L L S M S F S I L I I S P V S S A V G I A V G L T G L A S G A A N I G V S S C A M T L I V G T M 230

```

```

30 Query: 241  K A N K S G V P I A I A L G A M Q M M P N F L K H P V M A I P M L M A T V T S S L T V P L F K L V G T P A S S G F G L 300
      + N K G V P+A+ G A M K M+M P N++P++ I P+L+ V + L F L G T P A S+G G
30 Sbjct: 231  R V N K I G V P L A M F A G M M Q M L M P M I R Y P T I N I P L L I N G V C G V L A W L F N L Q G T P A S G P G F 290

```

```

35 Query: 301  V G A V G P I A S F E - A G A S M L ---- I V L S M L V I P F A V G F V S H K I C K D I L K Y K D I F V F E 353
      +G V G P I ++ A M+ I+ L+ V+ F ++ I D L K L Y++F+ E
35 Sbjct: 291  I G L V G P I N A Y R I M A Y T P M V R A G I L P L V Y P V L S F L A A Y L I D F L I V D R L K L Y R E L F I P E 348

```

There is also homology to SEQ ID 1280.

40 A related GBS gene <SEQ ID 8939> and protein <SEQ ID 8940> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1 Crend: 8
MoG: Discrim Score: -7.24
GvH: Signal Score (-7.5): -2.94

```

```

45 Possible site: 49
>>> Seems to have no N-terminal signal sequence
ALOM program count: 7 value: -9.82 threshold: 0.0
INTEGRAL    Likelihood = -9.82    Transmembrane  317 - 333 ( 304 - 335)
INTEGRAL    Likelihood = -7.64    Transmembrane  187 - 203 ( 183 - 217)
50 INTEGRAL    Likelihood = -6.37    Transmembrane  143 - 159 ( 136 - 161)
INTEGRAL    Likelihood = -5.26    Transmembrane   24 - 40 ( 18 - 44)
INTEGRAL    Likelihood = -2.34    Transmembrane  116 - 132 ( 115 - 136)
INTEGRAL    Likelihood = -2.13    Transmembrane   55 - 71 ( 55 - 71)
INTEGRAL    Likelihood = -0.96    Transmembrane  268 - 284 ( 268 - 284)
55 PERIPHERAL Likelihood = 0.69    205
modified ALOM score: 2.46

```

*** Reasoning Step: 3

```

60 ----- Final Results -----
      bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```


-2210-

Query: 5 IYDITTIIGGGPVGGLFAAFYAGLRGVSVKIIIESLSLGGQPAILYPEKKIYDIPGYVITG 64
 +YDITI+GGGP GLFAAFY G+R VKIIBS+ +LGGQ A LYPEK IYD+ G+P +
 Sbjet: 7 LYDITTIIGGGPVGGLFAAFYAGLRGVSVKIIIESLMPGLGGQIAALYPEKKIYDIPVAGFPKVA 66

Query: 65 RELIDKHIBQLERFDKSEIICKEEVLSPFK-VDDVFTTIQTDKQHLRAIVFACNGAF 123
 ++L++ Q E+F +I L++ V + K DD FTI+TDK+ H S+AI+ G GAP
 Sbjet: 67 QDLVNDLKRQABQNPITI--ALBQSVQNVTKETDPTITIKTKETHYSKAI IITAGAGAF 124

Query: 124 APRLLGLENEENYADNNLFYNVTKLEQFAGKHVVICGGGDSAVDWMANELOKIASVAIVH 183
 PR L +E + Y NL Y V L +AGK+V+I GGGDSAVDMA L+ +A +V ++H
 Sbjet: 125 QPRRLSVEGAKQYBGNLQYFVNDLNAYAGNVNLISGGGDSAVDWMALBPVAKNVTLIH 184

Query: 184 RRDAPRAHEHSVDILKASGVRLITFPYVIGLNGDSQRVSSLVQKVGDEVLPLDNL 243
 RRD FRAHEHSV++L+ S V ILT+ L+GD +++ + +Q+VKG D V L +D +I
 Sbjet: 185 RRDKFAHEHSVELLKSSVNIITPFAISELSGDEKIHVTTIQEVKGDAVETLDEV 244

Query: 244 VSPGFSTSNKLNRYNHLQYKRSSINVSLFFETQBGVYAIGDAANYPGKVELIATGYGEA 303
 V+PGF +S ++ W L+ ++SI V++ ET G+YA GD YPGKV+LIAT+GEA
 Sbjet: 245 VNFPGFVSLGPIKNGWLEIERNISVNTKMETNIPGIYAAGDICTYPGKVELIATGGEA 304

Query: 304 PVAINGAINITYIPDRDNRVHSTSL 328
 P A+N A +I P HSTSL
 Sbjet: 305 PTAVNAKAFIDPTARVFPGHSTSL 329

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6079> which encodes the amino acid sequence <SEQ ID 6080>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have an uncleavable N-term signal seq
 30 INTEGRAL Likelihood = -0.37 Transmembrane 8 - 24 (8 - 24)

----- Final Results -----
 bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CB15201 GB:Z99120 similar to thioredoxin reductase [Bacillus subtilis]
 Identities = 173/328 (52%), Positives = 223/328 (67%), Gaps = 4/328 (1%)

Query: 4 KAYDITTIIGGGPIGLFAAFYAGLRGVTVKIIIESLSLGGQPAILYPEKKIYDIPAYSLT 63
 K YDITTIIGGGP+GLF AFY G+R +VKIIBS+ +LGGQ + LYPEK IYD+ +P +
 Sbjet: 6 KVIDITTIIGGGPVGGLFTAFYAGLRGVSVKIIIESLPLGGQLSALYPEKKIYDIPVAGFPKIR 65

Query: 64 GVELTENLIKQLSRFEDRTITCLKEEVLTPFKVKGQ-PSIRTNKAHFSKAIITACNGA 122
 EL NL Q++P+ TICTL+ V +K G P + K K I GNGA
 Sbjet: 66 AQELINNLRQNAKFDQ--TICLEQAVESVEKQADGVFLVQMKKPTTLKRRCITACNGA 123

Query: 123 FAPRILGLESEENFADHMLFYNVHQLQDQFAGQKVVICGGGDSAVDMALEADIASVTVV 182
 F PR L L E+ E + NL Y V L +FAG++V I GGGDSAVDMA L E IA+ V+++
 Sbjet: 124 FKPRKLELENAEQYBGNLHYFVDDLQKFAGRRAVILGGGDSAVDMALEADIAKGVIII 183

Query: 183 HRRDAFRAHEHSVELLKASTVNLITFPYVFKALGIGNLAERKIVQKRVDEVLKLESL 242
 HRRD FRAHEHSVE L AS VN+ITP+VP L G + E+LV+++VK D L E+D L
 Sbjet: 184 HRRDKFRAHEHSVENLHASKVNLITFPVPAELIGEDI- BQLVLEEVGDKRKELEIDL 242

Query: 243 IVSPGFSTSNKLNRYNHLQYKRSSITVSLPQTSQGTI FALGDAAYNGKVELIATGYGE 302
 IV++GF +S +KNW LD ++SI V +T+ EG FA GD Y GKV+LID+GGE
 Sbjet: 243 IVNYGFVSLGPIKNGWLEIERNISVNTKMETNIBGFFAAGDICTYBGNVLIASGGE 302

Query: 303 APTAVNAQINITYIPDRDNRVHSTSLID 330
 APTAVN A Y+ P +HSTSL +
 Sbjet: 303 APTAVNAKAYMDPKARVQLHSTSLFE 330

65 An alignment of the GAS and GBS proteins is shown below.

-2211-

Identities = 242/324 (74%), Positives = 279/324 (85%)

Query: 6 YDITIIGGGPVLGFAAFYAGLRGVSKIIISLSLGGQPAILYPKKIYDIPGYVITGR 65
 YDITI+GGGP+GLFAAFYAGLRGV+VKIISLSLGGQPAILYPKKIYDIP YP +TG
 5 Sbjct: 6 YDITIIGGGPVLGFAAFYAGLRGVTKIISLSLGGQPAILYPKKIYDIPAYPSLTVG 65

Query: 66 SLIDKHIEQLERFKDSIRICLKEEVLSPKVVDDVPTITQDKDQHLRAIVFACNGAPAP 125
 EL + I+QL RF+D ICLKEEVL+P+KV F+I+T+K +H S+AI+ ACNGAPAP
 10 Sbjct: 66 ELTENLIQLSRFEDRTTCLICEEVLTPDKVKGGFSIRTNKASHPSKAIILACNGAPAP 125

Query: 126 RLGLLENEENYADNNILFYNYVTKLEQFAGKHVVICGGSDSAVDWNALELDAESVTVVHR 185
 R LGLLE+ENK+AD+HLFYNY +L+QFAG+ VVICGGSDSAVDWA L+ LA SV +VHR
 15 Sbjct: 126 RLGLLESENFAHNLFLFYVHOLDQFAGQKVVICGGSDSAVDWNALELDAESVTVVHR 185

Query: 186 DAFRAHEHSDVILKASGVRLITFPYVPIGLNKDSQKVSSLVVQKVGDEVLLEPLDNLIVS 245
 DAFRAHEHSDV+LKAS V +LTFYVP L G LV+QKVK DEV+EL LD+LIVS
 20 Sbjct: 186 DAFRAHEHSEVLKASIVNLITFPYVPAKLGKIGNLAEKLVIQKVKDEVLLELLOSLIVS 245

Query: 246 PGFSTSNHQLRYNLDYKRSSINVSSLFETTQEGVYAIGDAANYPGKVLATGYGSAFV 305
 PGFSTSNHQL+ NLDYKRSSI VS LF+T+QEG+YAIGDAA Y GRV+LATG+GSAF
 20 Sbjct: 246 PGFSTSNHQLNLDYKRSSITVSFLPQSGBIQFAGDAAAYNGKVDLIATGPGSAFT 305

Query: 306 AINQAINIYIPDRDNRVHSTSLI 329
 A+INQAINIYIPDRDNRVHSTSLI
 25 Sbjct: 306 AVINQAINIYIPDRDNRVHSTSLI 329

SEQ ID 6078 (GBS178) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 5; MW 37.4kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 8; MW 62.4kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1962

A DNA sequence (GBSx2071) was identified in *S. agalactiae* <SEQ ID 6081> which encodes the amino acid sequence <SEQ ID 6082>. This protein is predicted to be tRNA methyltransferase (trmD). Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.1496 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:BAH06198 GB:AP001515 tRNA methyltransferase [Bacillus halodurans]
 Identities = 144/246 (58%), Positives = 186/246 (75%), Gaps = 6/246 (2%)

Query: 2 MKIDILTLFPFMPAPLEHS-IVGKAKERGLLEINYNHFNRENAR-KSRHVDDPEYGGQGM 59
 MKID LTLFPFMP + HS I+ +A+REG + NPKR +R K + VDD PYGGG GM
 50 Sbjct: 1 MKIDFLTLFPFMPQVLHSSILKQAGRGVSPRVNPFREYSNKHKKVDVDPYGGQGM 60

Query: 60 ILRAQPIFDITDKIDAQKA--RVILLDFNGRTFDQFAEELSKEDELI FICGHYEGYDE 116
 +L QP+PD ++ + + RVIL+ P G TF Q AKEL++ + LI +CGHYEGYDE
 55 Sbjct: 61 VLSPOQLFDAVEDLTKSSSTPRVIMCPQGETPTQKAEELAQABHILLICGHYEGYDE 120

Query: 117 RIKS-LVTDKVSLSGDVLTGSELAMTMVDATVRLIPEVIGKRTSHQDSSFGSLLEYQ 175
 RI+S LVTDK+S+GD+VLTGSEL AM + D+ RL+P V+G ETS Q DSFS+GLLEYQ
 Sbjct: 121 RIRSYLVTDKVSLSGDVLTGSELAMTVADSVTRLLPAVLGNETSQAQDSFSGLLEYQ 180

-2212-

Query: 176 YTRPYDYLGMVFPVVLMSGHHENIRKWRLEQSLRKTLERRPDLLENYAMTDEERLILEKI 235
 YTRP D+ G VPDVL+SGHH+NI +WR EQSL++TLERRPDLLE +T+EE+ +L+ I
 Sbjct: 181 YTRPADFRGKWKVDPVVLMSGHHENIRKWRLEQSLRKTLERRPDLLEGRKLTSEBQELLDGI 240

5 Query: 236 KTEIER 241
 + + E+
 Sbjct: 241 RKQCEK 246

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6083> which encodes the amino acid sequence <SEQ ID 6084>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2705 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 195/240 (81%), Positives = 224/240 (93%)

Query: 2 MKIDILTLFFENFAPLEHSIVGKAKRGLLEINYNHNFRENAEKSRHVDDFYG3G3QGMLL 61
 MKIDILTLFFENFAPLEHSIVGKAKRGLLEINYNHNFRENAEKSRHVDDFYG3G3QGMLL 60
 Sbjct: 1 MKIDILTLFFENFAPLEHSIVGKAKRGLLEINYNHNFRENAEKSRHVDDFYG3G3QGMLL 60

Query: 62 RAQPIFDITDKIDAKARVILLDEAGRTFDQDFABELSKEDSLIFIGHYBGYDERIKSL 121
 RAQPIFDIT++I+A+K R+ILLDEAG+ F Q +AEL+ E+SLIFIGHYBGYDERIK+L
 Sbjct: 61 RAQPIFDITQIEAKKPRIILLDEAGKPFITQAYAEELAEELSLIFIGHYBGYDERIKTL 120

Query: 122 VTDEVSLGDFVLVTGGELAAMTMVDATVRLIPVIGKETS HQDSSFSSGLLEYPQYTRPYD 181
 VTDE+SLGDFVLVTGGELAAMTMVDATVRLIP+V+GKE+SHQDSSFSSGLLEYPQYTRPYD
 Sbjct: 121 VTDEISLGDFVLVTGGELAAMTMVDATVRLIPQVIGKESHQDSSFSSGLLEYPQYTRPYD 180

Query: 182 YLGNTVPDVLMSGHHENIRKWRLEQSLRKTLERRPDLLENYAMTDEERLILEKIKTEIER 241
 Y GNTVPDVLMSGHHE IR WRLE+SL+KT RRPDLLE+Y ++EER +L+KIK +++
 Sbjct: 181 YRGNTVPDVLMSGHHENIRKWRLEQSLRKTLERRPDLLEHNFSEERLLDKIKREALDQ 240

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1963

A DNA sequence (GBSx2072) was identified in *S.galactiae* <SEQ ID 6085> which encodes the amino acid sequence <SEQ ID 6086>. This protein is predicted to be 16S rRNA processing protein. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.71 Transmembrane 32 - 48 (32 - 52)

----- Final Results -----
 bacterial membrane --- Certainty=0.2084 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9381> which encodes amino acid sequence <SEQ ID 9382> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB13475 GB:299112 similar to hypothetical proteins [Bacillus subtilis]

-2213-

Identities = 88/174 (50%), Positives = 128/174 (72%), Gaps = 1/174 (0%)

Query: 54 VIMEYFNVGKIVNTQGLQGBMRVLSVDFVEERFKKGQVLAFLDEKNQFVMDIASHRK 113
 +T +FNVGKIVNT G+GB+RV+S TDF KRR+K G L DF + +++ +HR
 Sbjet: 1 MTKRWFNVGKIVNTGIGIKGEVRVISTDFABERYKPGNTLYLFMDGRNEPVEIVMTHRL 60

Query: 114 QKNFDIIFKGMHYHINDIKYKGFYLKVAEDQLSDLKDGEPFYHHEIGLDVYBGE-ELIG 172
 K F +++FK ++N++E+ K +KV E++L +L +GEFY+HEIG +V+ E ELIG
 Sbjet: 61 HKQFHLQDFKRNQANBYRELKNAIIVPEBRLGELNAGEFYFHEIGCEVFTBGLIG 120

Query: 173 KIKEILQPGANDVWVVERHGKRDLLLPYIPPVLEVDLSNQRVQVLEMEGLDDE 226
 K+KEIL PGANDVW+ R GK+D L+FYI VV +D+ ++++ELMEGL DE
 Sbjet: 121 KVKELITPGANDVWVIGRKGKDALIPYIESVVKHIDVREKKITELMEGLDDE 174

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6087> which encodes the amino acid sequence <SEQ ID 6088>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2787 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 133/172 (77%), Positives = 153/172 (88%)

Query: 56 MEYFNVGKIVNTQGLQGBMRVLSVDFVEERFKKGQVLAFLDEKNQFVMDIASHRKQK 115
 MEYFNVGKIVNTQGLQGBMRVLSV+DF EERFKKG LALFD+K++FV ++ I SHRKQK
 Sbjet: 1 MEYFNVGKIVNTQGLQGBMRVLSVDFARRRFKGSLALFDKDRFQVITVSHRKQK 60

Query: 116 NFDIIFKGMHYHINDIKYKGFYLKVAEDQLSDLKDGEPFYHHEIGLDVYBGEELIGKIK 175
 +FDIIFK MYHIN IKYK+G+TKV++D DL++GEFYH+IIG+ VYE + LiG +K
 Sbjet: 61 HFDIIFKGMHYHINAIRKYGTYLKVSKINQDLQBGEPFYHQIIGMAVTEKDVLIHGVK 120

Query: 176 EILQPGANDVWVVERHGKRDLLLPYIPPVLEVDLSNQRVQVLEMEGLDDED 227
 EILQPGANDVW+V+R GKRDLLLPYIPPVVL VD+ N+RV VEIMEGLDDED
 Sbjet: 121 EILQPGANDVWVIRKQGRDRLLPYIPPVVLNDVDPNKRVDVLEMEGLDDED 172

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1964

A DNA sequence (GBSx2073) was identified in *S.galactiae* <SEQ ID 6089> which encodes the amino acid sequence <SEQ ID 6090>. This protein is predicted to be similar to E. coli ykfc (11). Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3488 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9909> which encodes amino acid sequence <SEQ ID 9910> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-2214-

>GP:AAC38715 GB:AF030367 maturase-related protein [Streptococcus pneumoniae]
Identities = 366/425 (86%), Positives = 396/425 (93%)

5 Query: 12 MSLLDKILSRNNLEAYQVKSNGSGAGINVTIEQMDYLHQNWRETKQLIKERSYK 71
MS+LLDKILSR NNLEAY QVKSNGSGAGI+G+TIE+MD+YL QNWR TK+LIK+R YK
Sbjct: 1 MSKLLDKILSRNNLEAYQVKSNGSGAGIDGMTIEEMDNYLRQNWRLTKELIKQRIYK 60

10 Query: 72 QPVL+VEIPKP+GG+R LG+PT MDRMIQQAIVQV+SP+CE HFS+ SYGRRPNRSCE A 131
QPVL+VEIPKP+GG+R LG+PT MDRMIQQAIVQV+SP+CE HFS+ SYGRRPNRSCE A 120
Sbjct: 61 QPVLKVRIPKP+GGIRQLG+PTVMDRMIQQAIVQVMSPTCEPHFSDTSYGRPNRSCEKA 120

15 Query: 132 IVQLLEYINDGYEWIVDIDLEKFPDTPQDRIMSLVINII+DGTSTSLIRKYLHSGVIN 191
I++LLSYINDGYEWIVDIDLEKFPDTPQDRIMSLVINII+DGTSTSLIRKYLHSGVIN 191
Sbjct: 121 IMKLLLEYINDGYEWIVDIDLEKFPDTPQDRIMSLVINIIEGDTSTSLIRKYLHSGVIN 180

20 Query: 192 GQRHKTAVGTPQGGNLSFLLSNIMLNELDKLEKRLRFVRVYADDCVITVGESEAAAKRVM 251
GQR+KTAVGTPQGGNLSFLLSNIMLNELDK LEKRLRFVRVYADDCVITVGESEAAAKRVM 251
Sbjct: 181 GQRYKTAVGTPQGGNLSFLLSNIMLNELDKLEKRLRFVRVYADDCVITVGESEAAAKRVM 240

25 Query: 252 HVSYSYIEKRLGLKVINMTKTIKIVRPNKLYLGFPGFWKSPKRWKCRPHQDSVQSPKRRKLK 311
+SVS +IEKRLGLKVINMTKTI RP +LKLYLGFPGFWKSPKRWKCRPHQDSVQSPKRRKLK 311
Sbjct: 241 YVSRFPIEKRLGLKVINMTKTIKIVRPNKLYLGFPGFWKSSDGWKSPPHQDSVRRPKLKLK 300

30 Query: 312 LTVRKNSIDLTRIRBLNWIRGWINYPFSLGNMKSIMTQIDERLRTIRIRIWNQWKKKA 371
LT RKNSIDL RIE+LN IRGWINYPFSLGNMKSIMTQIDERLRTIRIRIWNQWKKKA 371
Sbjct: 301 LTVRKNSIDLTRIRBLNWIRGWINYPFSLGNMKSIVASIDERLRTIRIRIWNQWKKKS 360

35 Query: 372 RRLWGLLLKGVARWIADKVGSGWDHYQLVAQKSVLKRAISKPALAKRGLVSLCDYYLERH 431
+RLWGLLLKGV +WIADKVGSGWDHYQLVAQKSVLKRAISK L RKLVSCLDYYLERH 431
Sbjct: 361 RRLWGLLLKGVKWIADKVGSGWDHYQLVAQKSVLKRAISKPVLEKRLVSLCDYYLERH 420

Query: 432 ALKVS 436
ALKVS
Sbjct: 421 ALKVS 425

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1965

40 A DNA sequence (GBSx2074) was identified in *S.agalactiae* <SEQ ID 6091> which encodes the amino acid sequence <SEQ ID 6092>. Analysis of this protein sequence reveals the following:

Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -0.37 Transmembrane 7 - 23 (7 - 23)

45 ----- Final Results -----
bacterial membrane --- Certainty=0.1150 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 821> which encodes the amino acid sequence <SEQ ID 822>. Analysis of this protein sequence reveals the following:

Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -2.87 Transmembrane 1157 - 1173 (1157 - 1174)

55 ----- Final Results -----
bacterial membrane --- Certainty=0.2147 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

60

An alignment of the GAS and GBS proteins is shown below.

Identities = 1031/1064 (96%), Positives = 1042/1064 (97%)

5	Query: 1	MRKKQKLFPDKLAIALISTSILNNAQSDIKANTVTEDTPTAQEAPVEPQPIAVSESPSS 60
	Sbjct: 1	+RKKQKLEPDKLAIAL+STSILNNAQSDIKANTVTEDTPTAQEAVE PQP AVSEE+PSS 60
10	Query: 61	KETKTSQTPDSVGHETVADDANDLAPQAPAKTADTPATSKATIRDLNDPSHVKTQLRKAGK 120
	Sbjct: 61	KETKT QTF D ET+ADDANDLAPQAPAKTADTPATSKATIRDLNDPS VKTLQKAGK 120
15	Query: 121	GVGVVAVIDAGFDKNHRAWLTDKTKARYQSKENLEKAKKEHGITYGEWVNDKVAYYHD 180
	Sbjct: 121	G GIVVAVIDAGFDKNHRAWLTDKTKARYQSKLE+LEKAKKEHGITYGEWVNDKVAYYHD 180
20	Query: 181	YSKDGQIADVQEHGTHVSGILSNAPSEPMKEPYRLEGAMPEAQLLLMRVEIVNGLADYAR 240
	Sbjct: 181	YSKDGK AVDQEHGTHVSGILSNAPSE KEFYRLEGAMPEAQLLLMRVEIVNGLADYAR 240
25	Query: 241	NYAQAIRDVAVNLGAKVINMSPGNAALAYANLPDETCKAPDYAKSGKVSIVTSAGNDSFSG 300
	Sbjct: 241	NYAQA I DAVNLGAKVINMSPGNAALAYANLPDETCKAPDYAKSGKVSIVTSAGNDSFSG 300
30	Query: 301	GKRLFLADHPDYGVVGTFAAADSTLTVASYSFDPKQLTETATVKTDDQKEMFVLSTNR 360
	Sbjct: 301	GK RLFLADHPDYGVVGTFAAADSTLTVASYSFDPKQLTETATVKT D QKEMFVLSTNR 360
35	Query: 361	FEPNKAYDYAYANRGTKEDDFKDVKGKIALIERGIDFKKIANAKKAGAGVGLIYNDQD 420
	Sbjct: 361	FEPNKAYDYAYANRG KEDDFKDV+GKIALIERGIDFKKIANAKKAGAGVGLIYNDQD 420
40	Query: 421	KGFFIELEPNVDQMPAAFISRRDGLLLKNPQKITFNATPKV/LPTASGTLKSRFSSWGLT 480
	Sbjct: 421	KGFFIELEPNVDQMPAAFISR+DGLLL+NPQKITFNATPKV/LPTASGTLKSRFSSWGLT 480
45	Query: 481	ADGNIKPDIAAPGQDILSSVANNKYAKLSGTSMSAPLVAGIMGLLKQYETQYFDMTPSE 540
	Sbjct: 481	ADGNIKPDIAAPGQDILSSVANNKYAKLSGTSMSAPLVAGIMGLLKQYETQYFDMTPSE 540
50	Query: 541	RDLAKKVLMSATALYDEDEKAYFSPRQAGAGVDAKASAAATWVTDKINTSSKVKHLN 600
	Sbjct: 541	RDLAKKVLMSATALYDEDEKAYFSPRQAGAGVDAKASAAATWVTDKINTSSKVKHLN 600
55	Query: 601	NVSDKFEVTVTHNKSDKQBLYYQVTVQTDKVDGKHALAPKALYETSWQKITIPANSS 660
	Sbjct: 601	NVSDKFEVTVTHNKSDKQBLYYQVTVQTDKVDGK HALAPKALYETSWQKITIPANSS 660
60	Query: 661	KQVTFPIDASRFSKDLQAMKNGYPLBGPVRFKQDPTKEBLMSIPYIGPFGDQGNLSALE 720
	Sbjct: 661	KQVTFPIDVSQSKDLQAPMKNNGYPLBGPVRFKQDPTKEBLMSIPYIGPFGDQGNLSALE 720
65	Query: 721	KPIYDSKDGSSYTHEANSDAKDQLDQGLQPYALKNNPALTALTESNFWTIIKAVKGEVEN 780
	Sbjct: 721	KPIYDSKDGSSYTHEANSDAKDQLDQGLQPYALKNNPALTALTESNFWTIIKAVKGEVEN 780
70	Query: 781	IEDIESSEITETI PAGTFAKQDDSHYYIHRHANGKPYAALSPNGDGRDYVQFQPTFLR 840
	Sbjct: 781	IEDIESSEITETI PAGTFAKQDDSHYYIHRHANGKPYAALSPNGDGRDYVQFQPTFLR 840
75	Query: 841	NAKNLVAEVLDEKGNVWVTSVTEGVVKNYRNDLASTLGSTRFEKTRWDGKNDGKVVAN 900
	Sbjct: 841	NAKNLVAEVLDEKGNVWVTSVTEGVVKNYRNDLASTLGSTRFEKTRWDGKNDGKVVAN 900
80	Query: 901	GTYTYRVRYTPISSGAKGQHTDFVIVDNTIPEVATSATPSTEDSRLLTASKPKTSQPVY 960
	Sbjct: 901	GTYTYRVRYTPISSGAKGQHTDFVIVDNTIPEVATSATPSTEDSRLLTASKPKTSQPVY 960

-2216-

Query: 961 RERIAITYMDEDLPTTEYISPNEDGFTFLPERASTIMEGATVFLKMSDFTYVVEDMAGNIT 1020
 RERIAITYMDEDLPTTEYISPNEDGFTFLPERASTIMEGATVFLKMSDFTYVVEDMAGNIT
 5 Sbjct: 961 RERIAITYMDEDLPTTEYISPNEDGFTFLPERASTIMEGATVFLKMSDFTYVVEDMAGNIT 1020

Query: 1021 YTPVTKLLEGHSNKPEQDGSQDAPDKKPEAKPEQDGSQTPDKK 1064
 YTPVTKLLEGHSNKPEQDGSQDAPDKKPE KPEQDGSQ PDKK
 Sbjct: 1021 YTPVTKLLEGHSNKPEQDGSQDAPDKKPEKPEQDGSQAPDKK 1064

- 10 A related GBS gene <SEQ ID 8941> and protein <SEQ ID 8942> were also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 10
 MoG: Discrim Score: 5.69
 GVH: Signal Score (-7.5): -3.33
 15 Possible site: 25
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 1 value: -0.37 threshold: 0.0
 INTEGRAL Likelihood = -0.37 Transmembrane 7 - 23 (7 - 23)
 PERIPHERAL Likelihood = 2.81 508
 20 modified ALOM score: 0.57

*** Reasoning Step: 3

25 ----- Final Results -----
 bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 30 SEQ ID 8942 (GBS276) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 46 (lane 2; MW 123kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 5; MW 46.5kDa).

The GBS276-His fusion product was purified (Figure 206, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 296), which confirmed that the protein is immunoaccessible on GBS bacteria.

- 35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1966

A DNA sequence (GBSx2075) was identified in *S. agalactiae* <SEQ ID 6093> which encodes the amino acid sequence <SEQ ID 6094>. Analysis of this protein sequence reveals the following:

40 Possible site: 58
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 45 bacterial cytoplasm --- Certainty=0.4286(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

- 50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2217-

Example 1967

A DNA sequence (GBSx2076) was identified in *S.agalactiae* <SEQ ID 6095> which encodes the amino acid sequence <SEQ ID 6096>. Analysis of this protein sequence reveals the following:

```

Possible site: 30
5  >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -11.15    Transmembrane    19 - 35 ( 11 - 39)

----- Final Results -----
10      bacterial membrane --- Certainty=0.5458 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9911> which encodes amino acid sequence <SEQ ID 9912> was also identified.

15 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 6096 (GBS654) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 142 (lane 8 & 10; MW 51.2kDa + lane 9; MW 27kDa). Purified GBS654-GST is shown in Figure 245, lane 11.

20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1968

A DNA sequence (GBSx2077) was identified in *S.agalactiae* <SEQ ID 6097> which encodes the amino acid sequence <SEQ ID 6098>. Analysis of this protein sequence reveals the following:

```

25  Possible site: 14
    >>> Seems to have no N-terminal signal sequence

----- Final Results -----
30      bacterial cytoplasm --- Certainty=0.4174 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9913> which encodes amino acid sequence <SEQ ID 9914> was also identified.

35 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAF27324 GB:AF176424 unknwn [Lactococcus lactis]
Identities = 26/75 (34%), Positives = 45/75 (59%), Gaps = 4/75 (5%)

Query: 11 MAFEPKNSLTKVLKES-LDEKKKIPSEMNIRDPERTKYQYFTLPQSVRKKIDRLSKE 69
MAF+ + ++ VL SL + K E+ I E K Y FTL+PSV++ +++L+++
Sbjct: 1 MAFDVEDDKVKTVLNSSLAKSKVEL---PKLSENNKKYSFTLREPSVKEGLEKLAEK 57

Query: 70 KGYRASSSFINDFFK 84
+ Y++ S F+ND K
45 Sbjct: 58 QNYKNTSQFLINDLIK 72

```

No corresponding DNA sequence was identified in *S.pyogenes*.

-2218-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics:

Example 1969

A DNA sequence (GBSx2078) was identified in *S.agalactiae* <SEQ ID 6099> which encodes the amino acid sequence <SEQ ID 6100>. This protein is predicted to be ParA. Analysis of this protein sequence reveals the following:

Possible site: 45
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF27325 GB:AF178424 ParA [Lactococcus lactis]
Identities = 49/104 (47%), Positives = 72/104 (69%)

Query: 22 LSERLSEFTEAFDFKTRASYVTAKLFFLNMIKHNTNSSKELIRSLKNDKSVLWIMPHK 81
L ERL+ FK R D +TR +Y+TA +F+GN I+HNT SS+E + DK +AMIP K
Sbjct: 157 LIERLQNFDEVIDARTRETYITAIPIYFVGNRIRHNTSSREFSEKISQDKGTIAMIEK 216
Query: 82 ELFNKRSTLDKKGLSYMMSDKELYSRDSKFFKEIDFTFRKITDKL 125
ELFNKRSTLD L M DK+++ + F++++F F +IT+E+
Sbjct: 217 ELFNKRSTLDGVLVEMEKDDVFNHKKVFYEKLINFAPNEITNKI 260

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1970

A DNA sequence (GBSx2079) was identified in *S.agalactiae* <SEQ ID 6101> which encodes the amino acid sequence <SEQ ID 6102>. This protein is predicted to be transposase (orfA). Analysis of this protein sequence reveals the following:

Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2830 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1971

A DNA sequence (GBSx2080) was identified in *S.agalactiae* <SEQ ID 6103> which encodes the amino acid sequence <SEQ ID 6104>. This protein is predicted to be transposase (orfB). Analysis of this protein sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence

-2219-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2618 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP: CAB90834 GB: AJ250837 putative transposase [Streptococcus dysgalactiae]
 Identities = 242/259 (93%), Positives = 249/259 (95%)

Query: 1 MCRWLNMPHSYTYQAVESVSETEFEETIKRIFLDSSESYGSRKIKICINNNGITLSRRR 60
 MCRWLN+P SSYTY+AVE VSE E RE+IK IFL+S++RYGSRKIKICINNNGITLSRRR
 Sbjct: 1 MCRWLNIPRSTYTYKAVEPVSEAELESIKALFLESKARYGSRKIKICINNNGITLSRRR 60

15 Query: 61 IRRIMKRINLVSYYQKATFKPHSRGKNEAPIFNHLDRQFKPERPLQALVTDLTIVRVGNR 120
 IRRIMKRINLVSYYQKATFKPHSRGKNEAPIFNHLDRQFK ERPLQALVTDLTIVRVGNR
 Sbjct: 61 IRRIMKRINLVSYYQKATFKPHSRGKNEAPIFNHLDRQFKPERPLQALVTDLTIVRVGNR 120

20 Query: 121 WAYVCLIIDLYNREIIGLSLGHKTAELVKQIQSIYPALTKVMPHSDRXKEFDNQLID 180
 WAYVCLIIDLYNREIIGLSLGHKTAELVKQIQSIYP LTKVMPHSDR KEP+NQLID
 Sbjct: 121 WAYVCLIIDLYNREIIGLSLGHKTAELVKQIQSIYPPLTKVMPHSDRXKEFDNQLID 180

25 Query: 181 EILEAFGITRSLSQAGCPYDNAVAESTYRAFKEFVYQETPQLLEALATKDYVHWNY 240
 EILEAFGITRSLSQAGCPYDNAVAESTYRAFKEFVYQETPQ LLEALATK YVHWNY
 Sbjct: 181 EILEAFGITRSLSQAGCPYDNAVAESTYRAFKEFVYQETPQSLLEALATKAYVHWNY 240

Query: 241 HRIHGSINLYQTPMTKRLIA 259
 HRIHGSINLYQTPMTKRLIA
 Sbjct: 241 HRIHGSINLYQTPMTKRLIA 259

30

There is also homology to SEQ ID 32.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1972

35 A DNA sequence (GBSx2081) was identified in *S. agalactiae* <SEQ ID 6105> which encodes the amino acid sequence <SEQ ID 6106>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.3325 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1973

50 A DNA sequence (GBSx2082) was identified in *S. agalactiae* <SEQ ID 6107> which encodes the amino acid sequence <SEQ ID 6108>. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

-2220-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.4442 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9917> which encodes amino acid sequence <SEQ ID 9918> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AA044095 GB:AF115103 orf359 gp [Streptococcus thermophilus
 bacteriophage Sf121]
 Identities = 92/357 (25%), Positives = 162/357 (44%), Gaps = 33/357 (9%)

15 Query: 45 RKNQYKGTFTETMKEAYDELVRKIKYEFANKVSLNENYMTFENYMNKIYLRAKQK-VQSVT 103
 RK + F T EA ++ + + V+++ ++T +Y K+ YK+ V +T
 Sbjct: 24 RKPRTKGFRFKSEAIAAABMELKLDQNVNVD-E-DITLYDYP-KQCEVYKKTPTVSKIT 81

20 Query: 104 YKTALPHHKLFIQYFGLKPLKAITPRDCEAPRLHI IENYSENYAKNLMSRF----KACMG 159
 YK + + +FG K LK+IT + + ++ +Y++ +A++ RF KAC+
 Sbjct: 82 YKAYINSQRKIELFPFGDKCKLSITATYEQ---RVLSYAKTHAQDTVERPNVHVKACIE 137

25 Query: 160 YAERLGYISNMPCKALD---NPRGKHPTFPWYAEFQTFIKSFDLHDYBELQRFTAINWL 216
 A GYI CK +G+ ET F E++ I ++ + E + A+++
 Sbjct: 138 MAVHEGYIKRNFCKPFAKINAKNGRD IETKFLVEVEYERLI--YETSKHPEYASYAALYI 195

30 Query: 217 YMTGVRVSEGLSLQWEDIDFDDKPLKVHTTLEKDENGWYRKDQTKTPAGERLIELDDI 276
 TG+R +E L L +DI D L V+T + N + TKT + R I LDD
 Sbjct: 196 IAKTGRFACBGLTVDIDIKRTGMLSVNKTWYKNITGFM---PTKTKSSIRITPLDDE 252

35 Query: 277 TIEVLQVWRKNQFAMQDQDFIISFGDPFCKSTICRIIKRKAQQGVFVITGKGLRHS 336
 I + +Q D I+ + T+ +I+ R+ + LRH++A
 Sbjct: 253 FINFI-----DQLPPTDGRILPSELNNVAKTLRKIVGRE-----VRVHSLRHTA 299

Query: 337 SYLINVLKDDILYVARRMGHADKSTTINTYSHWFNALDKTVSEITTONIKSAGLDSI 393
 SYLI D++ V++ +GH + + TL Y+H E+I Q G +++
 Sbjct: 300 SYLI-AHDIDLISVSQVLGHENLNTILEVYAHQLQEQKSRNDEKIQMWTGCRNAL 355

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6109> which encodes the amino acid sequence <SEQ ID 6110>. Analysis of this protein sequence reveals the following:

40 Possible site: 61
 >> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.5549 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 111/127 (87%), Positives = 119/127 (93%)

50 Query: 242 LKVHTTLEKDENGWYRKQYTKTPAGERLIELDDITIEVLQVWRKNQFAMQDQDFIISRF 301
 LKVHTTLEKDENGWYRKQYTKTPAGERLIELDD+TI VL+ WR+NQ N DIDFIISRF
 Sbjct: 1 LKVHTTLEKDENGWYRKQYTKTPAGERLIELDDVTVVLENWRKNQVNTDQDFIISRF 60

55 Query: 302 GDPFCKSTICRIIKRKAQQGVFVITGKGLRHSASYLINVLKDDILYVARRMGHADKST 361
 G+PFCKSTICR+IK KAQ +GVFVITGKGLRHS+ASYLINVLKDDILYVA+ MGHADKST
 Sbjct: 61 GEPFCKSTICRVIKKAQSIGVFPVITGKGLRHSASYLINVLKDDILYVAKCMGHADKST 120

60 Query: 362 TINTYSH 368
 TINTYSH
 Sbjct: 121 TINTYSH 127

-2221-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1974

A DNA sequence (GBSx2083) was identified in *S.galactiae* <SEQ ID 6111> which encodes the amino acid sequence <SEQ ID 6112>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3299(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1975

A DNA sequence (GBSx2084) was identified in *S.galactiae* <SEQ ID 6113> which encodes the amino acid sequence <SEQ ID 6114>. This protein is predicted to be repressor protein-related protein. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2721(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9919> which encodes amino acid sequence <SEQ ID 9920> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC98432 GB:L29324 repressor protein [Streptococcus pneumoniae]
Identities = 38/65 (58%), Positives = 52/65 (79%), Gaps = 1/65 (1%)

Query: 2 MYRRLEDLREDNDPTQKYVARK-LSPTHSAYSKIIRGERILSADVLIKLNLNLYNVSTDYL 60
      M +R+EDLRED+D+Q+YVA+ L+ T SAYSK+E G R++S D +IKL++ YNVS DYL
Sbjct: 1 MLKRIEDLREDDDLTOEYVAKTILNCTRBAYSQMSGSTRILISIDDLIKLADFPNVSLDYL 60

Query: 61 LGQTD 65
      +G+ D
Sbjct: 61 VGRVD 65
```

There is also homology to SEQ ID 582.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2222-

Example 1976

A DNA sequence (GBSx2085) was identified in *S. agalactiae* <SEQ ID 6115> which encodes the amino acid sequence <SEQ ID 6116>. This protein is predicted to be relaxase. Analysis of this protein sequence reveals the following:

5 Possible site: 40
 >>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3160 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AAC98434 GB:L29324 relaxase [Streptococcus pneumoniae]
 Identities = 223/417 (53%), Positives = 310/417 (73%), Gaps = 5/417 (1%)

20 Query: 1 MVTIGHYAVHGKKYRQLIKLYLLDPKTRNLNLSDPGMSNYLDFPDYVELVWYQXNKL 60
 MVTIGHYAVHGGK YR +LLKYIL+P KT+NL+L+SDPGH NYLDFP Y ELVWY +NFL
 Sbjct: 1 MVTIGHYAVHGGK YR KLKLYLNPSKTNLTLVSDPGHNYLDFPSSYKELVWYNDNKL 60

25 Query: 61 SNDQLYDSRFDROKKQKKIHAHIIQSFSPEDKLSPREINRIGYETIKELGGQYKFIV 120
 SND LV+ R DRQE Q+KIH+HHIIQSPSP+D L+PE+INRIGYE KEL GG+++FIV
 Sbjct: 61 SNDTLVSEPHRDROQVWQKIHSHHIIQSFSPDDHLTPQELNRIGYEAAKELTGGPRFIV 120

30 Query: 121 ATHVDQDHCHNHIIINSINSQSKKLWMDYALERNLQMSDRISKVAGAKIIPPKRYSHR 180
 ATHVD+ H HNHII+NSI+ S KKK WDY E NL+H+SDR+SK+AGAKII RYSHR
 Sbjct: 122 ATHVDGKHIIHNSIILNSIDQNSDKKFLWMDYAEHNLRWVSDRLSKAGAKII-ENRYSHR 179

35 Query: 181 DYEVYRSNHHKYLKQRLFFPLMEHSIDFNDPMQAEQLNVKIDPSRKHSRPFMTDRNNKQ 240
 YEVYR++N+KYE+QOR++FL+E+S +F D +KA+ L++KIDF KH +FWTD NNRQ
 Sbjct: 180 QYEVYRTNIVKYEIKQRVVFLIENSKNFEDLKKKAKALHLKIDFRKHVTFYMDNNKQ 239

40 Query: 241 VIQSDKLNKREPYSEYFQRYPAKKICIELILEFLLRSNSPDDLVEKARILGLELKSKK 300
 V++ K++++PY++ YF++ F +++I ILEFLL + ++L+++A +GL++ K+K
 Sbjct: 240 VVRDSKLSRKQPYNETYPEKKFVQREIINILEFLLPKMONNELIQAQVFGDKIIPK 299

45 Query: 301 TIDFVLSDGKSCISIPNKSRLKKNLYDTTYFDSYFKEHDVPEVLHNEVKIEFEKFPETQQ 360
 + F DG I + + L K NLY +YF YF + VL N + + + + +
 Sbjct: 300 HVLFPF-DG---IKLAEQLVSNLYSVSYFQDYFINNKNETFVLNHNILVLAENEEKIK 355

50 Query: 361 LSEILTVSEITEAYETYTKTRDAVHFEFEVITEBQIEKIVLDGLFVKVMGIGQBEL 417
 E+ + E + ++Y+ +K RDAVHEFEVE+ QIE+V G++KV GI ++ L
 Sbjct: 356 EKELPSSEMWKSYQDFKTRRDVHFEFEVELNLNQIEVVEHGYIYKVGIDIKDL 412

45 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6117> which encodes the amino acid sequence <SEQ ID 6118>. Analysis of this protein sequence reveals the following:

50 Possible site: 20
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3114 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

55 An alignment of the GAS and GBS proteins is shown below.

 Identities = 103/218 (47%), Positives = 170/218 (77%)

60 Query: 393 EEQIEKIVLDGLFVKVWVGIGQBELIPNHQNLNLEQENKKQYQVPIRRTSYFYIYHKE 452
 E QIE+++ + ++KV + Q GLIFLN+QL+I ++EN K+Y+V+IKET+ +PY+KE
 Sbjct: 2 KHQIERLAIEDIYIKVSPSVKQGLIPNPNYQLDIRKENHKKYKVIYIKRTAQFFIYHKE 61

-2223-

Query: 453 DSENNRPMKGRDLIRQLTFDNKSLPYKRRISLVSLQKQIESINILNLTINIQNSFLLEKD 512
 SE+NR+M+G +LI QLT D+KS+P +RR ++ +L++KIEEI+L+ +NK + ++KD
 Sbjct: 62 ASLANRYMRGHELICQLTNDKSIKPKRRROTITDLKKKIEISILLIELDTENKPYQDIDK 121

5 Query: 513 ELVGDIAQLDIELTNLQDENFTLNKMAEYVVNLQSDNQDTKQLAKYECSSKNLSQNVITG 572
 ++V D+AQLD+ +T LQD LNK+AEV++NL +++ + ++L+Y+ +KGNL+ + I
 Sbjct: 122 DIVKMAQLDITTEIQLDHTAHLNKVAVLNLNNNDIENRLARYDYAKNLTAAIKIE 181

10 Query: 573 QIESIRMIQNLQDNKIEREYNAVRKLDQYVVRVLMNDK 610
 ++E EIE QN+L+ I+EYE VR+L+++ +L+ K
 Sbjct: 182 EYEKEITFSQRELNLISIDEYEVYLRLEKFGELSDSK 219

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1977

A DNA sequence (GBSx2086) was identified in *S. agalactiae* <SEQ ID 6119> which encodes the amino acid sequence <SEQ ID 6120>. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.4006 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC98436 GB:L29324 unknown [Streptococcus pneumoniae]
 Identities = 53/115 (46%), Positives = 77/115 (66%), Gaps = 2/115 (1%)

30 Query: 5 VREIRKVENFSISEYQIQIFMEQGVYQFSPFARGKLLKIDHOPQSLSEWIKYQLQHK 64
 +R IRK+ + E +QI + M ++G + FS F R LL D Q +Q+E+W + QK
 Sbjct: 5 IRSIRKQFLITETEEKQILDLMREKGDONFSDFLRKSLLLSGQ--KQMEKWFNLAKTKQ 62

35 Query: 65 VEQIYRDVHEILVAKLSQSVTMEHLLEILTCIKDLAKIEVITPLSYSPDKQYM 119
 +EQI RDVHE+ ++AK + VT EH+ I+LTCI++L+KE+E T PLS F +KYM
 Sbjct: 63 LEQISRDVHEVFIARTKNHVTHEVSVILLTCIQELIKEVEKTPGLSEDFCNKYM 117

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1978

A DNA sequence (GBSx2087) was identified in *S. agalactiae* <SEQ ID 6121> which encodes the amino acid sequence <SEQ ID 6122>. This protein is predicted to be TnpA. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.2935 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC82523 GB:AF027768 TnpA [Serratia marcescens]

-2224-

Identities = 176/413 (42%), Positives = 243/413 (58%), Gaps = 18/413 (4%)

Query: 26 MFPKVEAVGPPERPCECGFD-KLYKHSSRNQLIMDLPIRLKRVGLHNNRRYKCREOGST 84
 M P+V+ V P C ECG + + R+ DLPI KRV L + RRY CR C +T
 Sbjct: 1 MHPQVD-VPDPLACEBCGVQGEFVRPGKRDPYRDLPINGKRVTLMVVRRYTCRACKT 59

Query: 85 IS-----VDEKRSMTKRLKLSIQEQSMKTFVEVARSVGVDKTIIRNVFKYVALKERE 138
 VD R MT RL + +++S + + VA G+DEKT+R++F R
 Sbjct: 60 FRPQLPENVDGFR-MTLRLHEYVEKESFNHPYTFVAQTGLDEKTVRDI PNARAEPLGRW 118

Query: 139 YQFETPKWLGIIDEIHIIRPRVLATNIERRKTIYDIKWNRKETVQRLSEISDRTYIEYV 198
 ++FETP+ LGIDE+++ +R R +LNIIE RT+ D+ R ++ V L ++ DR +E V
 Sbjct: 119 HRFTPRILGIDELAYLNKRYRCILATNIERRKTLDDLATRRQDVVINYIMKLKDRQKVEIV 178

Query: 199 TMDMWPYKDAVNTILPQAKVVVDKPHVVRMANQALDNVRKSLKAHMSKERRTLNRRF 258
 +MDMWP PY+ AV +LPQA++VVDKPHVVRMAN AL+ VRK L+ + + RTL +R
 Sbjct: 179 SMDMWPYRAAVKAVLPQARIVVDKPHVVRMANDALERVKGLKELKPSQSRTLKDRK 238

Query: 259 ILLKRKHDLNERESFLLOTWLGNLPAKAEYELKEEFTWINDTPDDBGHLYSQMHRHC 318
 ILLKR H++++RE +++TW G P L AYE KE FY IWD + +W
 Sbjct: 239 ILLKRAHEVSDRERLIMETWTFAPQLLAAYEHKERYGIWDAITRLQAEALDEW-IAT 297

Query: 319 MSSNSKDAYKDLVPAVDNHHVEI PNYF--DKRLTNAYTESINSIIRQVRMGRGYSFDAL 376
 + K+ + DLVRAV NN E YF D +TNAYTESIN + + R GRGYSF+ +
 Sbjct: 298 IPIQGEKVDSDLVRAVGWREETMTYFETDMPVNTAYTESINRLAKDKNRBGRGYSFEVM 357

Query: 377 RAKILFNEKLHKRKRPRFNSSAFNKAMLYDTFNWYEVMDHDITDNLGVDFSTL 429
 RA++L+ K HKK+ P S P K + Y + D N GVD ST+
 Sbjct: 358 RARMLATTK-HKKKAPTAKVSPFYKTI-----GYGLPDFAEELNYGVDFSTI 404

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1979

- 35 A DNA sequence (GBSx2088) was identified in *S.galactiae* <SEQ ID 6123> which encodes the amino acid sequence <SEQ ID 6124>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2115 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA70224 GB:Y09024 mercuric reductase [Bacillus cereus]
 Identities = 412/546 (75%), Positives = 484/546 (88%)

Query: 1 MNKPKVNISGMTCTGCCEKIVESALEKIGAKNISSYRGRGAVFELPDDIEVSAIKAIDE 60
 M K++V+ + GMTCTGCE+HY ALE +GA IE +RGRGAVFELP+ +V+H KAI +
 Sbjct: 1 MKKQYVDVQGMTCTGCCEHVAVALENMGATGIEVDGRGAVFELPINALGVETAKLAISD 60

Query: 61 ANYQAGEIEEVSSLENVALINEDNYDLITGSGAAFPSSAKAIEYGAQVMIERGTVGG 120
 A YQ G+ EV S R V L NK +YD +ITGG AAFSSAI+A++YGAQV MIERGTVGG
 Sbjct: 61 AKYQGGKAEIEVQSQEMVQLGNBGGDYDIIITGSGAAPPSSAIEAVIKYGAQVMIERGTVGG 120

Query: 121 TCVNIGCVPSKTLIRAGEINHLKDNPFILQTSAGEVDLASLITQDKLIVSELNRQKYM 180
 TCVNIGCVPSKTLIRAGEINHL+K+NPF+GL TSAGEVDLA LI QK++L+HLRN KY+
 Sbjct: 121 TCVNIGCVPSKTLIRAGEINHLAKNNPFVGLTSSAGEVDLAPLIKQKNELVTELRNSKYV 180

-2225-

Query: 181 DLDEYNFDLIRGEAKFVDASTVEVNGTKLSAKRFLIATGASPSLQDISGLEKMDYLTST 240
 DLD+Y F+LI+GEAKFVD TVEVNG +SAKRFLIATGASP+ P I GL +DYLTST
 Sbjct: 181 DLDDYGFELIRGEAKFVDKRTVEVNGAPISAKRFLIATGASPAKPNIPGINVDYLTST 240

Query: 241 TLLELKKIKPRLTIVIGSGYIGMELGQLFHHLGSEITLMQRSERLIKEYDPEISESVKAL 300
 +LLELKK+KRL VIGSGYIGMELGQLFHHLGSE+TL+QSRKRLIKRYDPEISESVK+L
 Sbjct: 241 SLLELKKVPRLTIVIGSGYIGMELGQLFHHLGSEVTLTQSRERLKEVDPEISESVKSL 300

Query: 301 IBQGINLVKGTATPERVEQSGEIKRVYTVNGSRVIESDQLAVATGRKPNTDSNLASAAG 360
 +BQGINLVKGT+ER+EQ+G+IK+V+V VNG + +IB+DQLVATGR PNT +LNL AAG
 Sbjct: 301 VBQGINLVKGTATYERIBQGNIDIKRVHEVNGKRLIIBDQLVATGRPTNTATNLRAAG 360

Query: 361 VETGKNNEILINDFGQTSNKKIYAAGDVTLSGPQFVYVAYEQGIITONAGGLNKKIDLS 420
 VE G EI+I+D+ +T+N +IYAAGDVTLSGPQFVYVAY+QG+ NAIGGLNKK+L
 Sbjct: 361 VEIGSRGEIITIDYSRITNIRIYAAGDVTLSGPQFVYVAYQGVAAFPNAIGGLNKKINLE 420

Query: 421 VPEAVTFNTPNATVATGLTEQAKQKNGYDVKTSVLEFLDAVPRAIVNRETTGVFKLVADAET 480
 VVP VTFT P +ATVGLTE+QAKE GY+VKTSVLEFLDAVPRA+VNRETTGVFKLVAD++T
 Sbjct: 421 VVPGVTFTAPALATVATGLTEQAKQKNGYVKTSLVLEFLDAVPRAVNRETTGVFKLVADSKT 480

Query: 481 LKVLGVHIVSENAGDVIYAASLAVKFGLTIEDTLFTLAPYLTVABGLKVALTFDRIK 540
 +KVLG H+V+ENAGDVIYA+LAVKFGLT+D+ ETLAPYLTVABGLK ALTFDRIK
 Sbjct: 481 MKVLGAHVVAENAGDVIYAATLAVKFGLTVDIRETLAPYLTVABGLKALATFDRIK 540

Query: 541 LSCAG 546
 LSCAG
 Sbjct: 541 LSCAG 546

30 There is also homology to SEQ ID 1820.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1980

A DNA sequence (GBSx2089) was identified in *S.galactiae* <SEQ ID 6125> which encodes the amino acid sequence <SEQ ID 6126>. This protein is predicted to be regulatory protein. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4529 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA83973 (G:AF138877 mercury resistance operon negative
 regulator MerR [Bacillus sp. RC607]
 Identities = 83/129 (64%), Positives = 104/129 (80%)

Query: 1 MYRISEPADKQGVNKETIRYERKNLLQBPHTKTRAGYIYSDDVKRVQFIKRIQEPGF 60
 M+RI E ADKQGVNKETIRYER L+ EP RTE GYR+YS V R+ FIKR+QE GF
 Sbjct: 1 MKPRIGELADKQGVNKETIRYERKGLIPEPHTKTRAGYIYSDDVKRVQFIKRIQEPGF 60

Query: 61 SLSEIYKLLGVVDKDEVRQGMPEFVSKQKQKQVQIQEDLKRITMDDLKQRCFDEKIL 120
 +L+EL KLLGVVD+DE +CDK+PF K +++Q++IEDLKRIE ML DLK+RCF+ R +
 Sbjct: 61 TINEIDKLLGVVDREAKRCMYDFTILKIEDIQKIEDLKRIERMINDLKRCFENEDI 120

Query: 121 HSCPIIETL 129
 + CPIIETL
 Sbjct: 121 YSCPIIETL 129

-2226-

There is also homology to SEQ ID 1712.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 1981

A DNA sequence (GBSx2090) was identified in *S.agalactiae* <SEQ ID 6127> which encodes the amino acid sequence <SEQ ID 6128>. Analysis of this protein sequence reveals the following:

```

Possible site: 44
>>> Seems to have no N-terminal signal sequence
10  INTEGRAL    Likelihood = -7.86    Transmembrane    80 - 96 ( 78 - 100)

----- Final Results -----
      bacterial membrane --- Certainty=0.4142 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
15      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8943> and protein <SEQ ID 8944> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1    Crend: 8
McG: Discrim Score:    -13.52
25  GVH: Signal Score (-7.5): -6.14
      Possible site: 44
>>> Seems to have no N-terminal signal sequence
ALOM program    count: 1 value: -7.86 threshold: 0.0
      INTEGRAL    Likelihood = -7.86    Transmembrane    80 - 96 ( 78 - 100)
30  PERIPHERAL    Likelihood = 1.80    136
      modified ALOM score: 2.07

*** Reasoning Step: 3

35  ----- Final Results -----
      bacterial membrane --- Certainty=0.4142 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

40 The protein has homology with the following sequences in the databases:

```

ORF02021(439 - 666 of 1080)
GP[451734]gb|AAA18975.1||U05143(9 - 46 of 46) envelope glycoprotein {Simian
immunodeficiency virus} GP[451744]gb|AAA18980.1||U05148 envelope glycoprotein {Simian
immunodeficiency virus}
45 %Match = 3.2
%Identity = 38.5 %Similarity = 64.1
Matches = 15 Mismatches = 13 Conservative Sub.s = 10

336      366      396      426      456      486      516      546
50  RIFPQFGKCCDYNNVNGVPLSRINLEHYLTGGVLYFVVYSKDVSPVTVTYASLTPKVIKQNLPSADKKRIKKKEDIFL
      :||| : |||:|:|:|:|
      WGLTGNAGTTPTATTTTTTFRVVENVNESN-----
      10      20      30

```

-2227-

```

576      606      636      666      696      726      756      786
LPWGAIIAKLLILPYPALQTSYKSRPCLRRSSRLKLTQIPPSIVIKVGNINMKSITAFLOVKAYILPCLAGXGPARDIMVW
      |||: || |||
-----PCTIDNSCAGLECEP
      40

```

SEQ ID 8944 (GBS415) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 3; MW 21.2kDa).

Example 1982

A DNA sequence (GBSx2092) was identified in *S.agalactiae* <SEQ ID 6129> which encodes the amino acid sequence <SEQ ID 6130>. Analysis of this protein sequence reveals the following:

```

Possible site: 14
>>> Seems to have no N-terminal signal sequence

```

```

----- Final Results -----

```

```

bacterial cytoplasm --- Certainty=0.3402(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1983

A DNA sequence (GBSx2093) was identified in *S.agalactiae* <SEQ ID 6131> which encodes the amino acid sequence <SEQ ID 6132>. This protein is predicted to be ATPase. Analysis of this protein sequence reveals the following:

```

Possible site: 27
>>> Seems to have no N-terminal signal sequence

```

```

INTEGRAL Likelihood =-10.08 Transmembrane 324 - 340 ( 317 - 343)
INTEGRAL Likelihood = -5.73 Transmembrane 662 - 678 ( 660 - 690)
INTEGRAL Likelihood = -5.41 Transmembrane 350 - 366 ( 346 - 378)
INTEGRAL Likelihood = -3.40 Transmembrane 94 - 110 ( 93 - 110)
INTEGRAL Likelihood = -2.87 Transmembrane 681 - 697 ( 680 - 699)
INTEGRAL Likelihood = -1.38 Transmembrane 148 - 164 ( 148 - 164)

```

```

----- Final Results -----

```

```

bacterial membrane --- Certainty=0.5034(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAA22856 GB:M90750 cadmium-efflux ATPase [Bacillus firmus]
Identities = 486/725 (67%), Positives = 584/725 (80%), Gaps = 18/725 (2%)

```

```

Query: 1 MSRGKAKQSEKEMKAYRVQGFCTCNCAAFENNVKELDGVQDAKVFHGASKVYVGGTTII 60
MS KA SE+EMKAYRVQGFCTNCA FE NVK+L GV+DAKVNFGASK+ V G TI
Sbjct: 1 MSQKAITSEQEMKAYRVQGFCTCANCAKFEKNVQGLSGVREDKVNFGASKIAVYNGATTI 60

```

```

Query: 61 EELEKAGAPENLKIRDEHKQRVGGE-----PPWKQKNIKVYISALLLVSWFL 109
EELEKAGAPENLK+ EK R + PP+K K + +Y S LL+ +
Sbjct: 61 EELEKAGAPENLKVTPKSAQASQEVQEDTKEDKVPFYK-KHSTLLYAS-LLITFGYIS 118

```

```

Query: 110 GEQYGBEHVLPETGYASILIGGYSLFIEGLKNLRRLNFDNMFLMTIAIGAAIIGWGE 169
GER+++ T+ + AS+ IGG SLF GL+NL R FDM TMT+R+IG AIIGW E

```

-2228-

Sbjct: 119 SYVNGEENIVTLLFLASMFIOGLSLFKVGLQMLRPFSDMKTLMTAVVIGGAIGGEAR 178

Query: 170 GATVVLFAISEALERYSMDKARQSIRESMDIAPKEALIRRGNEEMIHVDEIQVDIMI 229
A VVILFAISEALER+SMD+ARQSI SIMDIAPKEAL++R +E+MIHVD+I VQDIMI

5 Sbjct: 179 VAIVVLFAISEALERFSDMRARQSIRESMDIAPKEALVKRNGQMIHVDIAVGDIMI 238

Query: 230 VKPGQKIAMDGIVVKGSTINQAAITGSSVPPVTKITINDEVFAGTINEEGILLEVKVTKRVE 289
VKPGQK+AMDG+VV G S +NQ AITGSSVPPV K ++EVFAGTINEEGILLEV++TK VE

10 Sbjct: 239 VKPGQKIAMDGIVVSGYSAVHQTAITGSSVPPVSKTVDNEVFAGTINEEGILLEVEITKLVE 298

Query: 290 DTTLSKIIHLVBEAQERAPSQAFVDKFAKYTTPAIVILALLIAVVPPL- PGDWSQMIY 348
DTT+SKIIHLVBEAQ ERAPSQAFVDKFAKYTTP I+I+A L+A+VPPL F G W MIY

15 Sbjct: 299 DTTISKIIHLVBEAQGERAPSQAFVDKFAKYTTPIIIMIAITLVAIVPPLFDGWSWEIY 358

Query: 349 QGLAVLVGCPALVSTPFAVVTAGNAAGKGVLIKGGIHLAAGHILKAIFDKTGLT 408
QGLAVLVGCPALV+STP+++V+AIGNAAK GVL+KGG++LE G LKAIFDKTGLT

20 Sbjct: 359 QGLAVLVGCPALVISTPISIVSAIGNAAKGVLVKGGVYLEEMGALKALAFDKTGLT 418

Query: 409 KGIPAVID--IVTYGRNENELITTSATKSGSQHPLASAIMRKABENGLKFEVITVEDPQ 466
KG+PAVID ++ NE EL++I +A+E SQHPLASAIM+KAER + ++V VEDP

25 Sbjct: 419 KGIPAVIDYNNINQINKEKLLSIITALEYRSQHPLASAIMKABEENITYSDVQVEDFS 478

Query: 467 SITGKGKAKINNEMYYGSGNLFEE-LHGSISSDKKEKIADMQTQGGKTMVLGTEKEIL 525
SITGKG+K +N YY+GS LF+E L D ++ + +Q QGKT M++GTEKEIL

30 Sbjct: 479 SITGKGIGKIVNGTITTYIGSPILKELLINDFDKDLQNVTTLQNGKKTAMIGTEKEIL 538

Query: 526 SFIAVADEMRSSKEVIGKLNMMGI-ETVMLTGDNQRTATAIGKQGVSDIKADLLEPK 584
+ IAVADE+RESSKE++ KL+ +GI +T+MLTGDN+ TA AIG QVGVSDI+A+L+P+DK

35 Sbjct: 539 AVIAVADEVRESSKEILQLKHLGSLIKTKTITMLTDKNKTANAIAGGQGVSDIARELMPQDK 598

Query: 585 INFIFELREKHQSGVMVGQGVNDAPALAASTVGVMGAGGTDTALETADIALMSDDLKIL 644
L+FIK+LR ++ +V MVGDGVNDAPALAASTVG+AMGAGGTDTALETAD+ALM DDL KL

40 Sbjct: 599 LDFIKQLRSEYGNVAMVGQGVNDAPALAASTVGIAMGAGGTDTALETADVALMSDDLKIL 658

Query: 645 FYTTLKSRKALAIYKQNTFSLAIKLVALLVMPGWLTLWIAIPADMGATLVLNLSRL 704
P T+KLSRK L IIK NITP++ATK +A LLV+PGWLTLWIAI +DMGATLV LN LRL

45 Sbjct: 659 PSTVKLSRKTLNIIKANITFAIAIKFIASLNLVTPGWLTLWIAILSDMGATLVNLALNRL 718

Query: 705 LKIKK 709
+++KK

50 Sbjct: 719 MRVKE 723

There is also homology to SEQ ID 3506.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1984

A DNA sequence (GBSx2094) was identified in *S.agalactiae* <SEQ ID 6133> which encodes the amino acid sequence <SEQ ID 6134>. Analysis of this protein sequence reveals the following:

Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.0779 (Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-2229-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1985

A DNA sequence (GBSx2095) was identified in *S.agalactiae* <SEQ ID 6135> which encodes the amino acid sequence <SEQ ID 6136>. Analysis of this protein sequence reveals the following:

```

Possible site: 58
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -8.92    Transmembrane    123 - 139 ( 115 - 145)
    INTEGRAL    Likelihood = -6.74    Transmembrane    172 - 188 ( 167 - 190)
10    INTEGRAL    Likelihood = -1.61    Transmembrane    80 - 96 ( 80 - 96)

----- Final Results -----
          bacterial membrane --- Certainty=0.4567 (Affirmative) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
15          bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9923> which encodes amino acid sequence <SEQ ID 9924> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 4216.

A related GBS gene <SEQ ID 8945> and protein <SEQ ID 8946> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1    Crend: 10
McG: Discrim Score:    -6.41
GVH: Signal Score (-7.5): -2.23
    Possible site: 58
>>> Seems to have no N-terminal signal sequence
ALOM program    count: 3 value: -8.92 threshold: 0.0
    INTEGRAL    Likelihood = -8.92    Transmembrane    123 - 139 ( 115 - 145)
30    INTEGRAL    Likelihood = -6.74    Transmembrane    172 - 188 ( 167 - 190)
    INTEGRAL    Likelihood = -1.81    Transmembrane    80 - 96 ( 80 - 96)
    PERIPHERAL    Likelihood = 2.92    46
    modified ALOM score: 2.28

35    *** Reasoning Step: 3

----- Final Results -----
          bacterial membrane --- Certainty=0.4567 (Affirmative) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
40          bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1986

A DNA sequence (GBSx2096) was identified in *S.agalactiae* <SEQ ID 6137> which encodes the amino acid sequence <SEQ ID 6138>. This protein is predicted to be histidine rich P type ATPase (HRA-1) (copB). Analysis of this protein sequence reveals the following:

```

Possible site: 54
>>> Seems to have no N-terminal signal sequence
10    INTEGRAL    Likelihood = -13.37    Transmembrane    318 - 334 ( 307 - 345)
    INTEGRAL    Likelihood = -5.84    Transmembrane    347 - 363 ( 335 - 364)
    INTEGRAL    Likelihood = -5.15    Transmembrane    88 - 104 ( 86 - 112)

```

INTEGRAL	Likelihood = -5.04	Transmembrane	651 - 667 (649 - 669)
INTEGRAL	Likelihood = -4.30	Transmembrane	156 - 172 (155 - 173)
INTEGRAL	Likelihood = -4.30	Transmembrane	669 - 685 (668 - 690)
INTEGRAL	Likelihood = -3.03	Transmembrane	62 - 78 (60 - 80)

----- Final Results -----

```
bacterial membrane --- Certainty=0.6349(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA62113 GB:U16658 histidine rich P type ATPase [Escherichia coli]

Identities = 598/731 (81%), Positives = 651/731 (88%), Gaps = 36/731 (4%)

Query: 1 MRNKKIISHSHHNIGDIDHSKIDHNTMEHSOMDHS----- 36

Shict: 1 MRNKK+HSHSHHHNHGD++HSHKHSHNEMEHSSQMDHS
MRNKKOHSSSHSHHHNHGDMEHSHKHSHNEMEHSSQMDHSAMGHCSAMGGHARHHHGHGDMHSHYHD 60

© 2006 The Authors
Journal compilation © 2006 Blackwell Publishing Ltd

Query: 37 -----NMDSMDHGAMGGGAHHHHHGSEFKIFLKSPLGLIAILLITPMMDIQL 84
MD+SEMDEHGAMGGGAHHHHHGSEFK+IFLKSPLGLIAILLITP+M IOL

Subject: 61 H N E N K H S O M D H S K M D Y S E M D H G A M G G H A H H H H G S F K D I F L K S L P L G I A I L L I T P L M G I O L 12

Query: 85 PRQITFPYADVVAAVLATTILYIEGCKPEYMGAKDEFNASKARGMMSLITLGITVSVAYSIVY 14

PFQIIPPYADVAAVLATILYIPGGKPF MGAKEFNENK PGMMSLITLGITVSYAYSVY 11

PFQIIPPYADVAAVLATILYIPGGKPF MGAKEFNENK PGMMSLITLGITVSYAYSVY 1

QUESTION	145	31232VUDORSUMDEFEFFETI.YI.YM.I.GHNTMKALCEACDOKALAFIUDKDAHULR	2
----------	-----	--	---

[illegible]

Subject: 181 AVAARYVIGEPVMDFFFEFTTLILIMLLGHWIENKALGBAGNAQKALAEIVPRDAHVVLE 24

Query: 205 DDSIETFPVSELQIGDVIRVQAGENVPADGIIRGESRVNEALVTGESKPIEKKTGDEVI 26

shigt: 241 DDSIEIRFV++LQ+GD+IRVQAGENVPADG I RGESRVNEALVIGESKPIER GDEV1

[illegible]

query: 265 GGSTNGGGVLYVEIKQTGDQSFISQVQTLISQAQSQPSRAENVAQKVASWLFYIAVVAAL 32
GGSTNG GVLVVEIKQTGD+SFISQVQTLISQAQSQPSRAEN+AQKVA WLFYIAV+ AL

Subject: 301 GGSTNGDGVLYVEIKQIGDKSFISQVQTLISQAQSQPSRAENLAQKVAGWLFYIAVLAAAL 36

Query: 325 IALLIWTIIADLPTAVIFTVTALVLIACPHALGLAIPLVVSRSTSLGASRGLLVIONREALE 38

Sbjct: 361 IALVIMMVIADVPTAVIFVTTLVIACPHALGLAIPLVLTARSTSLGASRGLLVKDRDALE 42

LTT ADVMLDKTGTLTTFEFLVIV + +DKY+++EI LL+GIE GSSHPIAQSI+++

Sbjct: 421 LTTTRADVMVLDRKTGLTTGFEKVLDELFDNDKYTKDEIVALLSGIEGGSSHPFAQSIISY 48

Query: 445 ARAKGIKSVSFDSIEIVSGAGIEGEANGHHYQLISQKAYGKALRMDIPKGATLSILVENN 50

Sbjct: 481 AEQQGIRPVSFDSIDVMSGAGVEGQANGHRYQLISQKAYGRNLDMDIPKGATISVLVEND 54

Query: 505 EAIGAVALGDELKETSRLIEVLKKYGIEPLMATGDNREAAQGVAEVLGIQYQANQSPED 56

EAIGAVALGDELK TS++LI+ LKK I+P+MATGDNE+AAQG AE+LGI Y ANQSP+D

Table 1. Summary of the study design and sample characteristics

Query: 565 KYKLVESMKNQKNTVIMVGDGVNDAPSLALADVGIAIGAGTQVALDSADIILTQSDPGDI 62
KY+LVE +K + K VIMVGDGVNDAPSLALADVGIAIGAGTQVALDSADIILTQSDPGDI

Sp1ct: 601 KYELVEKLKAEGKKVIMVGDGVNDAPSLALADVGIAIGAGTOVALDSADIILTQYSPGDI 66

Query: 625 ESFIELANKTTTRKMKONTVMGAGYNEFAIPPIAGLIPIGITIGPAPGAVINSISTVIVA 68

SPIELA KTRKMK+NLVWGAGYNFIAIPAAIAG+LAPIGITL PA AVLMSLSSTVIVA

Subject: 001 ASFIELDGRINCPRESIDENTVWAGSINF IAF IAPSIUAFIGI INSFAVAVERBONGIVIA 78

Query: 685 INAMTELEPR 693
INAMTE KLERK

INAMTLKLEPK
Sbjct: 721 INAMTLKLEPK 731

-2231-

There is also homology to SEQ ID 3506.

A related GBS' gene <SEQ ID 8947> and protein <SEQ ID 8948> were also identified. Analysis of this protein sequence reveals the following:

```

5 Lipop: Possible site: -1  Crend: 7
  Mcg: Discrim Score: -19.12
  Gvm: Signal Score (-7.5): -3.71
    Possible site: 27
  >>> Seems to have no N-terminal signal sequence
  ALOM program count: 7 value: -13.37 threshold: 0.0
10 INTEGRAL Likelihood = -13.37 Transmembrane 291 - 307 ( 280 - 318)
  INTEGRAL Likelihood = -5.84 Transmembrane 320 - 336 ( 308 - 337)
  INTEGRAL Likelihood = -5.15 Transmembrane 61 - 77 ( 59 - 85)
  INTEGRAL Likelihood = -5.04 Transmembrane 624 - 640 ( 622 - 642)
15 INTEGRAL Likelihood = -4.30 Transmembrane 129 - 145 ( 128 - 146)
  INTEGRAL Likelihood = -4.30 Transmembrane 642 - 658 ( 641 - 663)
  INTEGRAL Likelihood = -3.03 Transmembrane 35 - 51 ( 33 - 53)
  PERIPHERAL Likelihood = 0.74 103
  modified ALOM score: 3.17

20 *** Reasoning Step: 3

  ---- Final Results ----
    bacterial membrane --- Certainty=0.6349(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

ORF02015(220 - 2304 of 2604)
EGAD|37454|38974(1 - 731 of 731) histidine rich P type ATPase (HRA-1) [Escherichia coli]
GP|643613|gb|AA62113.1|U|16658 histidine rich P type ATPase [Escherichia coli]
PTR|JC2464|JC2464 probable copper-transporting ATPase (EC 3.6.1.-) HRA-1 -
Enterobacteriaceae spp.
%Match = 67.4
%Identity = 85.9 %Similarity = 93.7
Matches = 598 Mismatches = 43 Conservative Sub.s = 54

162      192      222      252
PFRENTYM*CRMKP*MFKSL*YNKEELMNRNKKHSHSHHNGDI-----
|||||:
MRNKNKSHSHSHHNGDMEHSSKHDHNMESHSQNDHSAHMCAMGGEAHHHH
10      20      30      40      50

294      324      354      384      414      444      474      504
---DHSKCHDHNMEHSQNDHNSMDHSEMDHMGAGCHAAHHRRHSFKETFLKSLPLGLIAILLITPMMDIQLPQIIFPYADV
|||||:
GDMDSKCHDHNMEHSQNDHNSMDHSEMDHMGAGCHAAHHRRHSFKETFLKSLPLGLIAILLITPLMGIQLPQIIFPYADV
70      80      90      100      110      120      130

534      564      594      624      654      684      714      744
VAAVLATILYIPGGKPFYMAKADHFMASKAPGMSLITLGLITVSAYSVIYAARVYTGEPVMDFFFEPTTILLIMLHGM
|||||:
VAAVLATILYIPGGKPFYMAKADHFMASKAPGMSLITLGLITVSAYSVIYAARVYTGEPVMDFFFEPTTILLIMLHGM
150      160      170      180      190      200      210

774      804      834      864      894      924      954      984
IEMKALGEAGDAQKALAEVLPHDAVHVRDLSIETRPVSELQIGDVIRVGAGHVRADGIIIRKSRVNAELVTGESKPI
|||||:
IEMKALGEAGDAQKALAEVLPHDAVHVRDLSIETRPVADLVQSDLIRVQAGENVADCTIQRGSKVNAELVTGESKPI
230      240      250      260      270      280      290

1014      1044      1074      1104      1134      1164      1194      1224
EKKGDEGGSTGGGVLVVEIKOTGDSFISQVOTLISQAQSPSRANRYAKVASKLFIYIAVVALLIAITLTIAD
|||||:
EKKGDEGGSTGGGVLVVEIKOTGDSKSFISQVOTLISQAQSPSRANRLAKVASKLFIYIAVLLIALVITNVLAD

```

-2232-

	310	320	330	340	350	360	370	
	1254	1284	1314	1344	1374	1404	1434	1464
5	LPTAVIFVTITVALVIA	CPHALGLAI	PLVVS	RSTSLGASRGLL	VKNREALELIT	TKADVWVL	DKTGTIT	TGEPFKVL
	VTAVIFVTITVALVIA	CPHALGLAI	PLVVS	RSTSLGASRGLL	VKNREALELIT	TKADVWVL	DKTGTIT	TGEPFKVL
		390	400	410	420	430	440	450
10	1494	1524	1554	1584	1614	1644	1674	1704
	KYSEETGLLAGIEAG	SSHP	IAQSI	VNHAEAKG	IKSVSPDS	TRIVS	GAGIEG	RANGHHYQLISQ
	KYSEETGLLAGIEAG	SSHP	IAQSI	VNHAEAKG	IKSVSPDS	TRIVS	GAGIEG	RANGHHYQLISQ
		470	480	490	500	510	520	530
15	1734	1764	1794	1824	1854	1884	1914	1944
	TISVLVEND	EAIGAVAL	GDELKETS	RNLIEVL	KKYGI	EPLMAT	GDN	EAAQGV
	TISVLVEND	EAIGAVAL	GDELKETS	RNLIEVL	KKYGI	EPLMAT	GDN	EAAQGV
		550	560	570	580	590	600	610
20	1974	2004	2034	2064	2094	2124	2154	2184
	NKTVIHV	GDGVND	PSLALAD	VGLAIG	ATQV	LD	SD	ADIIIT
	NKTVIHV	GDGVND	PSLALAD	VGLAIG	ATQV	LD	SD	ADIIIT
		630	640	650	660	670	680	690
25	2214	2244	2274	2304	2334	2364	2394	2424
	AAGLLAP	IGITL	GPARG	AVLMS	LS	TV	VA	INAM
	AAGLLAP	IGITL	GPARG	AVLMS	LS	TV	VA	INAM
		710	720	730				
30	AAGLLAP	IGITL	GPARG	AVLMS	LS	TV	VA	INAM

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

35 Example 1987

A DNA sequence (GBSx2097) was identified in *Sagalactiae* <SEQ ID 6139> which encodes the amino acid sequence <SEQ ID 6140>. This protein is predicted to be CopA. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.2197	(Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000	(Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000	(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA40599 GB:XS7326 ORF-1 [Thiobacillus ferrooxidans]
 Identities = 26/65 (40%), Positives = 40/65 (61%), Gaps = 2/65 (3%)

Query: 1 MKQRIILL--DGVKCGACANTVQERFSAIEGVESVSDLATKKAVLESQTRIDTTETLNAAL 58
 M Q+I L G+ CA CR++V++ I G+S +V LAT +A + Q+ I TR L AA+
 Sbjct: 1 MSQKIFLRITGWTCAHCAHSVEKALLGHGIDSAQVSLATNQARVLQSSITPALLAAV 60

Query: 59 AETNY 63
 + Y
 Sbjct: 61 TQAGY 65

There is also homology to SEQ ID 3510.

-2233-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1988

A DNA sequence (GBSx2098) was identified in *S.agalactiae* <SEQ ID 6141> which encodes the amino acid sequence <SEQ ID 6142>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3220 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1989

A DNA sequence (GBSx2099) was identified in *S.agalactiae* <SEQ ID 6143> which encodes the amino acid sequence <SEQ ID 6144>. This protein is predicted to be heavy-metal transporting P-type ATPase (b0484). Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -4.09 Transmembrane 131 - 147 ( 130 - 150)

----- Final Results -----
bacterial membrane --- Certainty=0.2635 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB01764 GB:U42410 heavy-metal transporting P-type ATPase
[Proteus mirabilis]
Identities = 96/153 (64%), Positives = 123/153 (80%)

Query: 2 KAVKALRRRGVEVIMITGDNKRTAKAIKQVIGDSVLSEVLPEDKAEVKKLQEAQKQVA 61
+AKAL G+V MITGDNK TKAIAKQ+CID ++EVL+ K +K+L + G KVA
Sbjct: 649 EATKALHALGLKVMITGDNKRTAKAIKQVIGDITVAEVLPEDKVAALKQLSQKGDQVA 708

Query: 62 NVGSGINDAPALQANVGIAGSGTQVAIESADIVIMRIDLTAVLTITDLSHATLRNIQ 121
VGDGINDAPALQANVGIAGSGTQVAIESADIVIMRIDLTAVLTITDLSHATLRNIQ 121
Sbjct: 709 FVGDGINDAPALQANVGIAGSGTQVAIESADIVIMRIDLTAVLTITDLSHATLRNIQ 768

Query: 122 NLFWFAYNIVGIPVAMGLLYIFGGLMSFMLA 154
ELFW FAYN + IPVA G+LY G+L+SP+ A
Sbjct: 769 NLFWFAYNALLIPVAMGLLYIPYNGMLSPIFA 801
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3505> which encodes the amino acid sequence <SEQ ID 3506>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -10.83 Transmembrane 328 - 344 ( 314 - 348)
```

-2234-

5
10
15
20
25
30

INTEGRAL	Likelihood = -7.01	Transmembrane	354 - 370 (347 - 377)
INTEGRAL	Likelihood = -3.24	Transmembrane	101 - 117 (100 - 117)
INTEGRAL	Likelihood = -2.97	Transmembrane	165 - 181 (165 - 185)
INTEGRAL	Likelihood = -2.34	Transmembrane	665 - 681 (662 - 684)
INTEGRAL	Likelihood = -2.18	Transmembrane	67 - 83 (66 - 83)
INTEGRAL	Likelihood = -0.64	Transmembrane	491 - 507 (490 - 508)
INTEGRAL	Likelihood = -0.59	Transmembrane	691 - 707 (691 - 707)
INTEGRAL	Likelihood = -0.43	Transmembrane	140 - 156 (139 - 156)

----- Final Results -----
bacterial membrane --- Certainty=0.5331 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

15 An alignment of the GAS and GBS proteins is shown below.

Identities = 92/152 (60%), Positives = 123/152 (80%)

Query: 4 VKALRRFGVVEVIMITGDNKRTAKALAKQVGIDSVLSEVLPEDKAEVKKLQSAQKVMV 63
V+AL + G+ IM+TGD+ TAKATA QVGI V+S+VLP+ KA + L+ G+KVMV
20 Sbjct: 544 VEAHLQGLIHTLTGHDHATATASQVGTIDVISQVLPEQKAGVIADLRSGQRKVMV 603

Query: 64 GDGINDAPALAAQNVGIAGSGTDVAIESADIVLMRNDLTAVLTITDLSHATLRNKQL 123
GDGINDAPALA A++GIA+GSSTD+AIESAD++LN+ D+ ++ + LS T+R +K+NL
25 Sbjct: 604 GDGINDAPALADVIGIAMSGTDIAIESADVILNKPMDLDLVKANSRVTMRIVKENL 663

Query: 124 FWAFAYNLVQIPVANGLLIYFGGLMSPMLAG 155
FWAF YR++ IPVANGLL++FGG L++PMLAG
30 Sbjct: 664 FWAFIYNLMIPVANGLLHLFGGLINPMLAG 695

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1990

A DNA sequence (GBSx2100) was identified in *S.agalactiae* <SEQ ID 6145> which encodes the amino acid sequence <SEQ ID 6146>. This protein is predicted to be CopY. Analysis of this protein sequence
35 reveals the following:

Possible site: 23
>>> Seems to have no N-terminal signal sequence

40
----- Final Results -----
bacterial cytoplasm --- Certainty=0.2067 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:AGI0085 GB:AF296446 CopY [Streptococcus mutans]
Identities = 63/139 (45%), Positives = 96/139 (68%)

Query: 8 TSITDAWVKVMRVKANDLVTSKTVSVLKERMDTESTIKTLGRLVHKGVLINTQGR 67
TSI++AEWVEMRVVWA ++S ++I+L N+ STIKT++ RL EKG L +++GR
50 Sbjct: 2 TSISNARWVKVMRVKARWMTSSSEIIAILSRTYCSASTIKTLTKISSEKNTLTSGRQGR 61

Query: 68 KFIYTNIVKEKAVNDPARDIWNRIKCKGVGVIGSIIIEDEVLSPDDIDRIEKLSEIKS 127
K+IY++ I E+EA+ ++F+RIC K ++I ++E+ ++ DI++LE +L KK+
55 Sbjct: 62 KYIYSSILSEERLRQQVSEVPSRIVCKYKHALRIHVESETPMTLSIEKLEKALLSKKA 121

Query: 128 FAVEVDQCCTEGQCDCHS 146
AV EV C C QCC C+E
60 Sbjct: 122 NAVPEVKNCIVGQCSCYE 140

60 There is also homology to SEQ ID 3502.

-2236-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6151> which encodes the amino acid sequence <SEQ ID 6152>. Analysis of this protein sequence reveals the following:

```

Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
5  INTEGRAL    Likelihood = -13.16  Transmembrane  143 - 159 ( 135 - 165)
   INTEGRAL    Likelihood = -9.13   Transmembrane  49 - 65 ( 43 - 71)
   INTEGRAL    Likelihood = -7.17   Transmembrane  73 - 89 ( 72 - 94)
   INTEGRAL    Likelihood = -6.00   Transmembrane  13 - 29 ( 9 - 33)
10  INTEGRAL    Likelihood = -2.71   Transmembrane  180 - 196 ( 179 - 197)
   INTEGRAL    Likelihood = -0.59   Transmembrane  112 - 128 ( 109 - 128)

----- Final Results -----
          bacterial membrane --- Certainty=0.6265 (Affirmative) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
15         bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:AAF42284 GB:AE002544 cadmium resistance protein [Neisseria
meningitidis MCS8]
10  Identities = 201/208 (96%), Positives = 205/208 (97%)

Query: 1  MRCFMIGNVVTSIIYSGTAVDLLIIILMFAFKRKRDKIINIYLGQFLGSVSILLISLL 60
      Sbjct: 1  MRCFMIGNVVTSIIYSGTAVDLLIIILMFAFKRKRDKIINIYLGQFLGSVSILLISLL 60

25  Query: 61  FAFVLDYIPSEKILGLLGLIPILFLGLKVLVLLGSDGGRATAKESGLSKDNKNIPLVAMITF 120
      Sbjct: 61  FAFVLDYIPSEKILGLLGLIPILFLGLKVLVLLGSDGGRATAKESGLSKDNKNIPLVAMITF 120

30  Query: 121  ASCGADNIGVFPVFFPTTLANLANLIVALLTFVWMYLLVFSQAQLAQVPSVGTLEKYSRW 180
      Sbjct: 121  ASCGADNIGVFPVFFPTTLANLANLIVALLTFVWMYLLVFSQAQLAQVPSVGTLEKYSRW 180

35  Query: 181  FIADVYLGLGMVILIRNNSFMLWVVLG 208
      Sbjct: 181  FIADVYLGLGIYILVIRNNSFMLWVVLG 208

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 71/200 (35%), Positives = 130/200 (64%), Gaps = 4/200 (2%)

40  Query: 1  NSQTIIAIGVYISTSIDYLVLIILFAQLSQNKQKNHIYAGQYLGTGLVAGSLVRAVY 60
      Sbjct: 5  MRCNVVTSIIYSGTAVDLLIIILMFAFKRKRDKIINIYLGQFLGSVSILLISLLFAFV 64

45  Query: 61  VNFVPEAMVGLLGLIPILVIGIRFAIVGEGEHESEIERLEQSKANKLQFWITVTLTIA 120
      Sbjct: 65  LQVYIPSEKILGLLGLIPILFLGLKVLVLLGSDGGRATAKESGLSKDNKNIPLVAMITF 121

50  Query: 121  S-GQDNLGIYIPYFASLQNSQTLVLLVFAIGIIPCELSWVLSIPLISIEIKYQRII 179
      Sbjct: 122  SCGADNIGVFPVFFPTTLANLANLIVALLTFVWMYLLVFSQAQLAQVPSVGTLEKYSRW 181

55  Query: 180  VPLVPIGLGLNIMYSGTIE 199
      Sbjct: 182  IADVYLGLGMVILIRNNSF 201

```

SEQ ID 6150 (GBS174) was expressed in and purified from *E.coli*. The purified protein is shown in lane 7 of Figure 223.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2237-

Example 1993

A DNA sequence (GBSx2103) was identified in *S. agalactiae* <SEQ ID 6153> which encodes the amino acid sequence <SEQ ID 6154>. This protein is predicted to be Pgm. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.4324 (Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB96418 GB:AJ243290 phosphoglucomutase [Streptococcus thermophilus]
 Identities = 65/76 (85%), Positives = 71/76 (92%)

Query: 1 MTTYENLQKWLDFEQLPDYLRQELLSDMDEKTKEDAFYTNLEPGTAGMRGYIGAGTNRINI 60
 M+YTEN QKWLDF +LP YLR EL+SMDEKTKEDAFYTNLEPGTAGMRG IGAGTNRINI
 Sbjct: 1 MSYTYENQKWLDFEQLPDYLRQELLSDMDEKTKEDAFYTNLEPGTAGMRGLIGAGTNRINI 60

Query: 61 YVVRQATSEGLAKLIET 76
 YVVRQATSEGLA+LI++
 Sbjct: 61 YVVRQATSEGLAQLIDS 76

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6155> which encodes the amino acid sequence <SEQ ID 6156>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.4324 (Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 75/76 (98%), Positives = 75/76 (98%)

Query: 1 MTTYENLQKWLDFEQLPDYLRQELLSDMDEKTKEDAFYTNLEPGTAGMRGYIGAGTNRINI 60
 MTTYEN QKWLDFEQLPDYLRQELLSDMDEKTKEDAFYTNLEPGTAGMRGYIGAGTNRINI
 Sbjct: 1 MTTYENFQKWLDFEQLPDYLRQELLSDMDEKTKEDAFYTNLEPGTAGMRGYIGAGTNRINI 60

Query: 61 YVVRQATSEGLAKLIET 76
 YVVRQATSEGLAKLIET
 Sbjct: 61 YVVRQATSEGLAKLIET 76

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1994

A DNA sequence (GBSx2104) was identified in *S. agalactiae* <SEQ ID 6157> which encodes the amino acid sequence <SEQ ID 6158>. This protein is predicted to be a membrane protein. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -6.21	Transmembrane	94 - 110 (93 - 115)
INTEGRAL	Likelihood = -4.14	Transmembrane	172 - 188 (166 - 188)

-2238-

INTEGRAL Likelihood = -1.97 Transmembrane 130 - 146 (129 - 149)
 INTEGRAL Likelihood = -0.16 Transmembrane 62 - 78 (62 - 79)

----- Final Results -----

- 5 bacterial membrane --- Certainty=0.3484 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 10 >GP:CAA80247 GB:Z22520 membrane protein [Bacillus acidopullulyticus]
 Identities = 47/185 (25%), Positives = 80/185 (42%), Gaps = 23/185 (12%)
- Query: 1 MKKKKXSGNIATIAIFFAIMLVHFLSPFSPFWLVPKIPILMHIPVITASLAYGRIGA 60
 MKK +I I + A+ +++ T+MHIP II I GP +G
- 15 Sbjct: 1 MKKSLTVREIVAGVLGAVAILLGVTRLGYIPVPTAAGNATIMHIPAIIIGIMQGPVVGL 60
- Query: 61 TLGALMGGISVANGSIVLLPTSYLSPFVBNNGFYSLIALVPRILIGIIPFYVYKILHN 120
 +GA+ G S N+++ L F ++++PR+ IG++ + VY +
- 20 Sbjct: 61 IVGALIGISSFLNATVPL-----FKDPLVSILPRLPFGVAVMLVYIGIRR 105
- Query: 121 R---FGLAISGAIGSLTNTVFVLSGIFPESSTYNGNIKIMAGIISNSNAEMVIAAII 177
 + + +S IG+LTNT VL+ F + +A +N L E V+ I+
- Sbjct: 106 KSEYVAVGLSAFIOTLTNTALVLA--NAVFRHYLTAGVAVTVA---ITNGLPEAVVGTIV 160
- 25 Query: 178 VYLTV 182
 V
 Sbjct: 161 TLAVV 165

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6159> which encodes the amino acid sequence <SEQ ID 6160>. Analysis of this protein sequence reveals the following:

- 30 Possible site: 31
 >>> Seems to have an uncleavable N-term signal seq
- INTEGRAL Likelihood = -8.97 Transmembrane 18 - 34 (10 - 41)
 INTEGRAL Likelihood = -7.43 Transmembrane 170 - 186 (160 - 191)
 35 INTEGRAL Likelihood = -5.63 Transmembrane 96 - 112 (94 - 117)
 INTEGRAL Likelihood = -4.67 Transmembrane 140 - 156 (131 - 158)
 INTEGRAL Likelihood = -3.66 Transmembrane 64 - 80 (63 - 84)
 INTEGRAL Likelihood = -0.22 Transmembrane 39 - 55 (39 - 55)
- 40 ----- Final Results -----
 bacterial membrane --- Certainty=0.4588 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 45 The protein has homology with the following sequences in the databases:

- >GP:CAA80247 GB:Z22520 membrane protein [Bacillus acidopullulyticus]
 Identities = 47/193 (24%), Positives = 86/193 (44%), Gaps = 28/193 (14%)
- 50 Query: 8 RKSADISRIATIAIFFAIMLVHFLSPFVBNNGFYSLIALVPRILIGIIPFYVYKILPNQ 124
 +KS + I I + V + P+P T++HIP II ++ GP +G I
 Sbjct: 2 KKSLLTVREIVAGVLGAVAILLGVTRLGYIPVPTAAGNATIMHIPAIIIGIMQGPVVGLI 61
- Query: 65 LGGMLGILSVIINTIILLPTNYLSPFVBNNGFYSLIALVPRILIGIIPFYVYKILPNQ 124
 +G + GI S + T+ L F +++I+PR+ IG+ + Y I +
- 55 Sbjct: 62 VGAIFGISSFLNATVPL-----FKDPLVSILPRLPFGVAVMLVYIGIRR 106
- Query: 125 FGLIVSGI---IGSLTNTIFVLS-GIFIPATVFDGNIKALLTAISSNAIVEMISAI 180
 + G+ IG+LTNT VL+ +F + T + + +N + E ++ I+
- Sbjct: 107 SEYVAVGLSAFIOTLTNTALVLA VAVFRHYLT-----GVAVTITNGLPEAVVGTIV 160
- 60 Query: 181 TFFVLPTLSRLKR 193
 T ++ ++ R
 Sbjct: 161 TLAVVLAWKQIGR 173

-2240-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1995

- 5 A DNA sequence (GBSx2105) was identified in *S. agalactiae* <SEQ ID 6161> which encodes the amino acid sequence <SEQ ID 6162>. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have no N-terminal signal sequence (or aa 1-18)

- 10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.0165 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC44502 GB:U48885 DNA/pantothenate metabolism flavoprotein
[Streptococcus mutans]
Identities = 101/145 (69%), Positives = 122/145 (83%)

- 20 Query: 1 MIKRLTAVTGSISAYKAADLTSLQTKIGYDVHIINTQAATFPTPLTLQVLSKNPIHLD 60
M K+I LNV+GSI+AYKAADL+ QLTG+GY V++ NT AA +FI PLTLQVLSKNF++ +
Sbjct: 1 MTCKILLAVSGSISAYKAADLSHQLTKLGYHNVFWTNAKQFIPPLTLQVLSKNPVYSN 60
- 25 Query: 61 VMDEHNPKLIINHILAKRTDLFIVAPASANTIAHLAYGFADNIVTSVALAMPDETCKLIA 120
VM E +P++INHI LAK+ DLF++ PASANT+AHLA+GFADNIVTSVALA+P E PK A
Sbjct: 61 VMKEDDFQVINHILAKQADLFLPPASANTLAHLAHLGFADNIVTSVALALPDEVKFFFA 120
- Query: 121 PAMNTRKMYNHTITQRNIDILKKIGY 145
PAMNTRKY N ITQ NI +LKK GY
30 Sbjct: 121 PAMNTRKYENPITQSNITLKKFGY 145

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6163> which encodes the amino acid sequence <SEQ ID 6164>. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have no N-terminal signal sequence

- 35 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.0076 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
40 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 146/178 (82%), Positives = 155/178 (87%)

- 45 Query: 1 MIKRLTAVTGSISAYKAADLTSLQTKIGYDVHIINTQAATFPTPLTLQVLSKNPIHLD 60
M K I TLAV+GSISAYKAADLTSLQTKIGYDVHIINTQAAT+FITPLTLQVLSKN IHLD
Sbjct: 1 MTGHITLAVSGSISAYKAADLTSLQTKIGYDVHIINTQAATQFPTPLTLQVLSKNAIHLD 60
- 50 Query: 61 VMDEHNPKLIINHILAKRTDLFIVAPASANTIAHLAYGFADNIVTSVALAMPDETCKLIA 120
VMDEH+PK+INHILAKRTDLFIVAPASANTIAHLAYGFADN+VTSVALA+P TPCKLIA
Sbjct: 61 VMDEHDPKVINHILAKRTDLFIVAPASANTIAHLAYGFADNIVTSVALALPATTCKLIA 120
- Query: 121 PAMNTRKMYNHTITQRNIDILKKIGYQERPRISLLACGDTQSGALADISTILCKIQEV 178
PAMNTRKY N ITQ NI L IG+ EI P+ SLACGD G GALADI IL I +
55 Sbjct: 121 PAMNTRKYQNPITQENIKRLSTIGFTLPPKSSLLACGDKGPGALADIVILATITDT 178

SEQ ID 6162 (GBS236) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 5; MW 21.6kDa).

-2241-

Purified GBS236-GST is shown in Figure 208 (lane 6) and in Figure 225 (lanes 4-5).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1996

- 5 A DNA sequence (GBSx2106) was identified in *S. agalactiae* <SEQ ID 6165> which encodes the amino acid sequence <SEQ ID 6166>. This protein is predicted to be pantothenate metabolism flavoprotein homolog (dfr). Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence
 10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2325 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

15 A related GBS nucleic acid sequence <SEQ ID 9835> which encodes amino acid sequence <SEQ ID 9836> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:ANG39941 GB:AF301375 MTW1216 [Methanothermobacter wolfeii
 20 prophage psiM100]
 Identities = 71/229 (31%), Positives = 117/229 (51%), Gaps = 27/229 (11%)
 Query: 6 MKILITSGGTTEKIDTVRSITNHTATGTLGKIIAEKYLREBHQVTLVTTKNAVKPESATNL 65
 +++L+ GGT E ID VR ITN ++G +G +A + +G VTLV V + + L
 25 Sbjct: 172 LRLVLSLGGTLEPIDFVRVITNRSSGRMGLAVAREAYIQGAIVTLVA--GTVSVDPISQL 229
 Query: 66 STFEIEDVDGLIKTLKPLVKEHDILIHSMVSDYTPVYMADEFKVKSSDHLDTPLRKDNH 125
 T E + + + L+ EHD+ + + AVSD+ PVY
 30 Sbjct: 230 RTVRASTAHMAZAAVELIGRHDVFSAAAASVDFRPVYS----- 268
 Query: 126 EKGISSESEYQVFLKTKPKVISLWKWNPQITLVGPKLLVNVITKENLKVVARHSLIMNK 185
 E KISS+SE L LK PK+I + + NP+ +VGFK V++E L AR + +
 35 Sbjct: 269 EKKISSDSEI-TLRLKNPKIIRMAETNPFAIVGPKAEHGVSEKELIAAARKQIEDSV 327
 Query: 186 ATFILANDL-IDITSKHIAIYLLDHDNVYKATT--KEDIAQLIYEKVKK 231
 A ++AND+ ++ + ++ + V + T KE++A LI ++ K
 40 Sbjct: 328 ADMVVANDVSVEGPGSENNAITVSEGVTELPMTKKEBLAGLIIGRINK 376

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6167> which encodes the amino acid
 40 sequence <SEQ ID 6168>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence
 45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1737 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 142/230 (61%), Positives = 170/230 (73%)
 50 Query: 4 MAMKILITSGGTTEKIDTVRSITNHTATGTLGKIIAEKYLREBHQVTLVTTKNAVKPESAT 63
 M MK+ITSGSITE ID VR ITN+TG LCK+I E++L+ H VLTPTK A KP
 Sbjct: 1 MTMKLITSGGTTEPIDAVRGITNHSQQLGKILITERPLQYHHVDVLTPTKTKPLPNK 60

-2242-

Query: 64 NLSTFEIEDVDLSLIKTLKLVKSHDILHSMASVDYTPYIMADFRKVKSSDHLOTFLEKD 123
 L E+E V+ L+ LK V HDILHSMASVDYTPYIM D E+V +D+L+ FL +
 Sbjct: 61 RLRIIEVETVNDLAAALKDQVPHHDILHSMASVDYTPYIMDLBQVSGADNLNCLCEH 120

5 Query: 124 NHBKISSSESEYQVLELAKCTFKVISLVKKNPQITLWGFKLLVNVTKENLKFVARSGLIK 183
 N E KISS S+YQVLELAKCTFKVIS VK+WNF 1 LVGFKLLVNV +S L KVAR SL K
 Sbjct: 121 NSBPKISSASDYQVLELAKCTFKVISYVKNPNIKLWGFKLLVNVQRELLIKVARASLAK 180

10 Query: 184 NKATFILANLIDITSGHIIAYLLDHDNVYKATTKEDIAQLIYERVKCYD 233
 N A +ILANL+DI + H A L+ ++ V A TKE IA L+YE+ K+D
 Sbjct: 181 NHADYILANLVDIQTGMIKALLISNNEVASADTKRAIALDLLEYMTMGD 230

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 1997

A DNA sequence (GBSx2107) was identified in *S. agalactiae* <SEQ ID 6169> which encodes the amino acid sequence <SEQ ID 6170>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.22 Transmembrane 117 - 133 (117 - 133)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1086 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9833> which encodes amino acid sequence <SEQ ID 9834> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:BA07541 GB:AP001520 unknown conserved protein in B. subtilis
 [Bacillus halodurans]
 Identities = 94/221 (42%), Positives = 133/221 (59%), Gaps = 2/221 (0%)

35 Query: 52 AEKPIWTEVFLREINRSNQEIIHITWMTKTIVLGMIDRELPHLEAKHEIISRGYEPV 111
 A + F + + I + S L W TV+LG+ D LP + + + + +
 Sbjct: 27 ALQSFAYDOTLCTSIGKSPPTLRAMVHINTVVLISQDSRLPQIKAGIEALGQPHDVI 86

Query: 112 VNMFGGLAVADSGIILNFSLVIPDVFERKLGISDGYLIMVDFIRISDFVQPIEHFEVE 171
 VNM GGLAVV D GIIL SLV+ + E+ SI DGY +M + I S+V D + IE E+
 40 Sbjct: 87 VRNSGGLAVLDGGIILNLSLVKE -EKGFSIDGYELMYELICMPQDHRQIEAREIV 144

Query: 172 TSYCPGKFDLSINGKKFAGLAQRRIKNGIAVSIYLSVGDQKGRSQMISDFYKIGLGDTG 231
 SYCPG +DLSI+GKKFAG+ +QRRI+ G+AV IYL V G R+MI FY +
 45 Sbjct: 145 GSYCPGSDLSIDGKKFAGISQRRIKNGVAVQIYLCVSGQCAERAKMIRTFYDKAVAGQP 204

Query: 232 SPIAYPNVDPEIMANLSLDLCPMTFVEDVDRMLSLKQVG 272
 + YP + PE MA+LS+LL P V DV+ + L+L+Q G
 Sbjct: 205 TKFVYPRIKPEIMASISLLOQP+INVSVDLLKALMTLQHG 245

50 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6171> which encodes the amino acid sequence <SEQ ID 6172>. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -0.22 Transmembrane 95 - 111 (95 - 111)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1086 (Affirmative) < succ>

-2243-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

- 5 >GP:BAR07541 GB:AP001520 unknown conserved protein in *B. subtilis*
 [Bacillus halodurans]
 Identities = 97/228 (42%), Positives = 138/228 (59%), Gaps = 2/228 (0%)
- 10 Query: 30 ALSFPVWTEVFLKTIQNEPNQLILHIWMTITVLGLMDRLQFLYELAKTEIGNNGYVPV 89
 AL F + + + I + + L W TV+LG+ D +LP + + + +
 Sbjct: 27 ALQSFAYDDTLCTSIGKSQSPFTRLRWVHINTVVLGIQDSRLPQIKAGTEALKGQFQDVI 86
- Query: 90 TRNIGGLAVVADDGIILNFSVLIPDHPFSESISINAYLIMVDVIRESPSDYQRIEYHEIK 149
 RN GGLAVV D GIILN SLV+ + + SI + Y +M ++I F D+ ++IE EI
 15 Sbjct: 87 VRNSGGLAVVLDGIILN.SLVLKEE--KGPSIDDDGYRIAMYELICSMFQDRREKREIV 144
- Query: 150 NSYCPGNFDLSIAGRKFGAGIQRRIKIGIVVSIIYSLVCGDQVARGQLIKDFYEAGTQGEV 209
 SYCPG++DLSI G+KFGAGI+QRRI+ G+ V IYL V G A R ++I+ FY+ G+
 20 Sbjct: 145 GSYCPGSIYDLSIDGKKFAGISQRIRIGGVAVQIYLCVSGSGAERAKMRTFYDKAVAGQP 204
- Query: 210 TKVNYQIDPECMATLSSELLETPPTVAEVLRLRLTLRQLGFSLTEKS 257
 TK YP+I PE MA+LSELL P V++VL + +TL+Q G SL +S
 Sbjct: 205 TKFYPRIKPETMASISELLOQPINVSDVLKALMTLQHGASLILTES 252
- 25 An alignment of the GAS and GBS proteins is shown below.
 Identities = 155/275 (56%), Positives = 199/275 (72%), Gaps = 8/275 (2%)
- Query: 32 QDLAQLVPSIFKDYVIDAODAEKPFITWTEVFLIREINRSNOEIIILHIWMTITVLGLMDR 91
 +DLA LP+ ++ D A PF+TWTEVFL+ IN+ ++ILHIWMT+TVILGLMDR
 30 Sbjct: 10 RDLASLPIFIVYQDNKKVPGALSPFVWTEVFLKTIQNEPNQLILHIWMTITVLGLMDR 69
- Query: 92 ELPHLELAKKEIISRGYEFVVENFGJLAVVADGILNFSVLIPDVFERKLSISDGYLIMV 151
 +LP+ ELAK EI + GY PV RN GGLAVVAD+GILNFSVLIPD F +SIS+ YLIMV
 35 Sbjct: 70 QLPHYELAKTEIGNNGYVPVTRNIGGLAVVADDGIILNFSVLIPDHPFSESISINAYLIMV 129
- Query: 152 DFIRESFSDYQRIEYHEIKNFSYCPGNFDLSIAGRKFGAGIQRRIKIGIYVSLVCGD 211
 D IR FSD+YQ IE+ E++ SYCPG FDLST G+KFGAGIQRRIK GI VSIYSLVCGD
 Sbjct: 130 DVIRESPSDYQRIEYHEIKNFSYCPGNFDLSIAGRKFGAGIQRRIKIGIYVSLVCGD 189
- 40 Query: 212 QKGRSQMISDPYKIGLGDGTSPFIYVNDPEIMANSLDLLDCPWTEVDVIRMLSLKQV 271
 Q R Q+I DPF+ G + + YP +DPE MA LS+LL+ P TV +V++R+ ++L+Q+
 Sbjct: 190 QAARGQLKDFYEAGTQGEVTKVNYQIDPECMATLSSELLETPPTVAEVLRLRLTLRQL 249
- 45 Query: 272 GFN-----DRLIMIRPDLVAEPNRPQAKSMANK 300
 GF+ D+ L+ D V + R Q + + +G
 Sbjct: 250 GFSLTEKSPDQALLTNFDV--YERMQLEVVREKG 282

A related GBS gene <SEQ ID 8951> and protein <SEQ ID 8952> were also identified. Analysis of this protein sequence reveals the following:

- 50 Lipop: Possible site: -1 Crend: 10
 MoG: Discrim Score: -16.85
 GVH: Signal Score (-7.5): -5.07
 Possible site: 49
- 55 >>> Seems to have no N-terminal signal sequence
 ALON program count: 1 value: -0.22 threshold: 0.0
 INTEGRAL Likelihood = -0.22 Transmembrane 117 - 133 (117 - 133)
 PERIPHERAL Likelihood = 0.47 73
 modified ALON score: 0.54
- 60 *** Reasoning Step: 3
- Final Results -----
 bacterial membrane --- Certainty=0.1086 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-2244-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF01564 (451 - 1116 of 1518)

EGAD|13388|RS3758 (27 - 249 of 281) hypothetical 31.4 kd protein in pta 3' region {Bacillus subtilis}

OMNI|NT01BS4391 hypothetical protein SP|P39648|YWFL_BACSU HYPOTHETICAL 31.4 KDA PROTEIN IN PTA 3'REGION. GP|414014|emb|CNA51646.1||X73124 ipa-90d {Bacillus subtilis}

GP|2636300|emb|CAB15791.1||Z99123 alternate gene name: ipa-90d {Bacillus subtilis}

PIR|S39745|S39745 ywFL protein - Bacillus subtilis

%Match = 15.8

%Identity = 40.8 %Similarity = 61.0

Matches = 91 Mismatches = 82 Conservative Sub.s = 45

321 351 381 411 441 471 501 531

*WNLRETWGLSSDCDKINLAEPFRRMSDILLEWQDLAQLPVSI PKDYVTDAQDAEKPFIVTEVFIREINRSQEIILHI

MANQPIDLMMQPKNRVIDQSSLGPIFDAKQSFAMDDTLCMSVKGKGVSPATARS

10 20 30 40 50

561 591 621 651 681 711 738 768

WFMTKTVILMLRLPHLELAKKEIISRGVEPVVRNFGGLAVVADEGIILNFSLVIPDVFERK-LGISDGLYLMVDFIRS

WVHHDTIVLGIQDTRLPFIQDGLSILESGYRVIVRNSGGLAVVLDDGVNLISLIFED--ENKGGIDIDKGYEAMVELMR

70 80 90 100 110 120 130

798 828 858 888 918 972 996

IFSDFYQYIHFVEVETSCPGKFDLSINKKIFAGLAQRRIKNGIAVSIYLSVCGDQKG--RSQMISDFYKILGLD--TGS

MLRPYNAKTEAYRIEGSYCPGSDYLSINKKIFAGISQRRVGRGVAVQTYL--CADKSGSERADLIRRFVQALKDKQNDK

150 160 170 180 190 200

1026 1056 1086 1116 1146 1176 1206 1236

PIAYPMVDPEIMANLSLIDDCPTVEDVIDRMLISLQVGFNDRLLMRPDLVAEFNRFOAKSMANKGVSRDE*CPR*F

KGVPYIRPETMASISLILQKDISQDLMFALLTELKALSTHLYSAGLSIDSEMEFEKNILVRMARENKAVPG

220 230 240 250 260 270 280

SEQ ID 8952 (GBS390) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 73 (lane 7; MW 37kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 82 (lane 3; MW 62kDa).

GBS390-GST was purified as shown in Figure 216, lane 12.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1998

A DNA sequence (GBSx2108) was identified in *S. agalactiae* <SEQ ID 6173> which encodes the amino acid sequence <SEQ ID 6174>. This protein is predicted to be probable trimethylamine dehydrogenase (nema). Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2218 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA83700 GB:Z33015 similar to trimethylamine DH (Mycoplasma

-2245-

capricolum]
 Identities = 162/311 (52%), Positives = 219/311 (70%), Gaps = 1/311 (0%)

5 Query: 3 NVQGNLFRPLTLPLNGLSLENRFVLSFMVINSSTSEGFVTDGDIAYAVRRKASAPLQITGA 62
 N L F P L N G L E N R F V L S F M + T + G + T D + Y Y R R + S A P L Q I T G
 Sbjct: 2 NKYEKLEFPFYL-NGFKLENRFVLSFMVLSLTLATLDGKITDKEADYVRRSSASAPLQITGG 60

Query: 63 AYITEYQQLFEYGFVSVDKEDIPGLTKLAKAMKSKGAKAVLQUTHAGRFSSSHLLARHGV 122
 Y E+GQLFEYGS D+DIP LT+L + MK+ +LQL HAG+PS +L +GY+
 10 Sbjct: 61 VYDFEFGQLFEYGISAKSDDDIPSLTRLYQGMKTDSCNVILQAHAGKPSKTSLSKKGYL 120

Query: 123 YGSPFMQLQSPYPHQVKELTHKDLIRIIDEYVQATRRAIQAGFDGVEISSAQRLLIQTFP 182
 YGFS + +P R+V EL + I +II +Y AT R I+AGP+G+EIS AQRLLIQTFP
 Sbjct: 121 YGSPYKNNHTPIEHEVLELPEKIKQLIIQYKDATHRVIKAGFNGIEISSAQRLLIQTFP 180

15 Query: 183 STPSNRKDEYGPQTLNRCRLGLEVPKAVQKVIRESABSDFLGFRATPEETRGSQIGY 242
 S N+R DEY NR R LEV KA++VI + A +PI GFRATPEET G +GY
 Sbjct: 181 SQIINKRTDEYSATNFENSRFCLEVKALREVIDKYAPKNFIPGFRATPEETGIDLY 240

20 Query: 243 SIEEPMEFLKILAIQVLYLAIASWGHDFVNTIRSEGVYKQLVNQVIFEHFGORVPI 302
 +IE+P++ ++KI+ I ++ YLAIASWGH+ N +RS YGQLVNH+VI++ ++PI
 Sbjct: 241 TIEDFIQLVDKILIEIGKISYLAIASWGHDIYLNKVRSNVYKQGLVNHVIYDIYHKLPI 300

25 Query: 303 NRTGINSASAK 313
 +++GGIN+ +K
 Sbjct: 301 ISSGGINTVTK 311

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6175> which encodes the amino acid sequence <SEQ ID 6176>. Analysis of this protein sequence reveals the following:

30 Possible site: 35
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 35 bacterial cytoplasm --- Certainty=0.3055 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 265/390 (67%), Positives = 321/390 (81%)

40 Query: 8 LFRPLTLPLNGLSLENRFVLSFMVINSSTSEGFVTDGDIAYAVRRKASAPLQITGAAYITE 67
 L F P L T L P L N G L S L E N R F V L S F M V I N S S T + G + V T D D + Y A + R R A S A P L Q I T G A Y I T E
 Sbjct: 8 LFEPLTLPLNGSQLDNRFLSPFMVINSSTKDGVTQDDVSYALRRASAPLQITGAAYVDP 67

45 Query: 68 YQGLFEYGFVSVDKEDIPGLTKLAKAMKSKGAKAVLQUTHAGRFSSSHLLARHGVYGPSP 127
 YQGLFEYGFVSVDKD DI GL +LA+AMK+KGAKAVLQUTHAGRP+SH L ++G+VYGPSP
 Sbjct: 68 YQGLFEYGFVSVDKADISGLKELAQAMKAKGAKAVLQUTHAGRFSHALTYKGFVYGPSP 127

50 Query: 128 MQLQSPYPHQVKELTHKDLIRIIDEYVQATRRAIQAGFDGVEISSAQRLLIQTFPSTPSN 187
 MQL+SP PH+VK LT + I +I Y QATRRAIQAGFDGVE+SSAQRLLIQTFPSTPSN
 Sbjct: 128 MQLRSPQHEVKPLTQQQIEELIAAQAATRRAIQAGFDGVESSAQRLLIQTFPSTPSN 187

55 Query: 188 QRKDYGPQTLNRCRLGLEVPKAVQKVIRESABSDFLGFRATPEETRGSQIGYISREF 247
 +R D YG QTL NR +I L V +AVQ+VI++RA FI GFRATPEETRGA IGYSI+EF
 Sbjct: 188 KRTDSYGQTLNRRSKLTLAVLQAVQVQIKQAPDQFIPGFRATPEETRGNIDYSISDEF 247

60 Query: 248 MEFLKILAIQVLYLAIASWGHDFVNTIRSEGVYKQLVNQVIFEHFGORVPIAATGG 307
 ++ ++ +L +A++DYLAIASWG VFENT+RS G Y G+ VNQV+ ++ ++P+HAATGG
 Sbjct: 248 LQIMDWLVNVAKLIDYLAIASWGHDFVNTIRSPGYGRVNVQVRDYLRNKLPLVAATGG 307

65 Query: 308 INSASKVFALQAHMIGASTPLVDPEFLQIKAKCSQDINLRKVSLEGALIPKAF 367
 +N+ K EAL HA IG STP VVDPEF KIK C + I+LRI+ +DL+ LAIP+ASF
 Sbjct: 308 NMTDPKATEALAHADFIGVSTPVDPEFAHKIKGECESHLRLRPADLSLAIPQASP 367

Query: 368 KDIVPLADYGBSLPKAEAREVPRELRSNYRE 397

-2246-

KDIVPLMDYGESLPKE+R +FR L NY+E
 Sbjct: 368 KDIVPLMDYGESLPKESRTLPRLTINYKE 397

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1999

A DNA sequence (GBSx2109) was identified in *S.agalactiae* <SEQ ID 6177> which encodes the amino acid sequence <SEQ ID 6178>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3748 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA04594 GB:AP001510 unknown conserved protein [Bacillus halodurans]
 Identities = 121/333 (36%), Positives = 192/333 (57%), Gaps = 12/333 (3%)
 Query: 1 M KLSVLDYGLIDYKGTASDAIQETILLSQEAERLGYHQFWVAEHHGVKAFSINPELMIM 60
 M KLSVLD I YG A +A+++T L++ E LGYH+FWV+EHH + S+PE++I
 Sbjct: 1 M KLSVLDQSPIAYGSAKCALRQTELAKEVTEALGYHHRFVWSEHHDASTLAGSSPEVLIA 60
 Query: 61 HLANQTKSIKIGSGOIMPLHYSSFKLAETLKTLETCNHRVSIGLGNLGTVKVSNALRS 120
 HLA TK I++GSGG+M HYS++K+AR K L&E HP R+ +GUG + G + ++
 Sbjct: 61 HLAANTKKIRLSSGVMPLPHYSAYKVAENFKLLALHPGRIDVGLGRAPGMPFARMALQ 120
 Query: 121 LHK---AHDYEEVLEELKSWLIDESSSKEPL-----VQPTLSSFPDYLVLGSSQKSAYLAA 173
 K H Y +++ +L D+ + P + + PD+++LGS SA +AA
 Sbjct: 121 BGHEQNIHKYPLQKVDYIGYLDQLDPLDHRPHGLKATPLIDTVPDVWLLGSSQGSANVAA 180
 Query: 174 KLGLSFTPGVFPFMDKDP LTRAKKLSLTYHQFSEYYPNKSPLNMAVAFVADTSEAE 233
 + G GF F F++ + +A + Y F+ P VA FV+ ADT E+A+
 Sbjct: 181 ENGTGFAPA--HFINGEGVQAVE---SYRETPQPSALFDRPCTVAIFVICADTSEQAD 235
 Query: 234 NIAKTLDIWMLGNKDPNEFATPTTIEBNHYQLTTSQKAKIRSNRDMIVGDPKQVKESL 293
 IA +LD+ ++ ++ P+IE A Y +P ++A+I+ NR RMIVG PK V++ L
 Sbjct: 236 QIASSLDLSLIMLENOQLSKGTPTIESALSYPYSPFERARIRENRKRMIVGSPKAVRQOL 295
 Query: 294 DALVNASQAEILLIPLVPLGDQRKSLKLLSQ 326
 L A + EE+++ + + RI+S +LL +
 Sbjct: 296 VELARAYETEEVIVVTITHRPEDRIRSYELGE 328

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6179> which encodes the amino acid sequence <SEQ ID 6180>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.60 Transmembrane 212 - 228 (210 - 229)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2041 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 173/329 (52%), Positives = 241/329 (72%), Gaps = 1/329 (0%)
 Query: 1 M KLSVLDYGLIDYKGTASDAIQETILLSQEAERLGYHQFWVAEHHGVKAFSINPELMIM 60

-2247-

MK+S+LDYG+ID KY +A+ ET L+Q A++LG+H+FWVAEH+ AF+IS+PD++M
 Sbjct: 1 MKVSILOYGVIDREKTPQKALLETCTCLAQVADKLPHRPFVVAHHNIYAFALISSPELLMM 60

Query: 61 HLANQTSIKIGSGGIMPLHYSSFKLAETLKTLETCHPNKVSIGLGNLSLTVQVSNALRS 120
 HLA+ TK I+IGSGGIMPLHYSSFK+AR +TLE HPNR+ +G+GNSLGT V AL S
 Sbjct: 61 HLAHDTKIKIGSGGIMPLHYSSFKLAETWIMTLALHPNRIDLGISGSLTITLVQRLALS 120

Query: 121 LAKAHYRKEVLKELKSWLIDESSKPEL-VQPTTSSFPOLYVLGSGQKSAIYAAKILGLF 179
 +H Y +V+ EL +L + S P+ V P +P +L + +A LA +LGLG+
 Sbjct: 121 ICKDSYSQVVELYQYINPDHLSPLPIFVHPKNTYFQIWTLSNLSLETALQQLGLGY 180

Query: 180 TFGVFPFMDKDLTEAKKLSLYHQFREYTPMKSPNLMVAAPVVIADTSERENIAKTL 239
 TFG+FP++ KDP+TRAK++S+ Y F K P L++A P+V++DT B+AE +AK L
 Sbjct: 181 TFGIFPIYIPDPIFTRAKRVSAYHKAFRPSKLLKIPKILAVFIVLSUTDEKALAKPL 240

Query: 240 DIMMLGNKDPSFATFPPTIEANHYQLTPEQKAKIKSNRDMIVGDFQKESLDALVNA 299
 DIMMLG +DNEF T+P +EEA +Y LT +Q+ I +NR RM++G P VR+ LD L+ A
 Sbjct: 241 DIMMLGQQQDNFETKYTPDVERARNYHLEKQRELAANRSMVIGSPHYVKQLDRLEA 300

Query: 300 SQARELLLIPLVGLDQRIKSLKLSQLY 328
 QA+ELL IFLVP K +L+LL+ LY
 Sbjct: 301 CQDELLAIPLVPEFANRQRTLELLADLY 329

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 2000

A DNA sequence (GBSx2110) was identified in *Sagalactiae* <SEQ ID 6181> which encodes the amino
 acid sequence <SEQ ID 6182>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2384(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA781345 GB:AC007767 Identical to a glycine cleavage system
 H-protein precursor from Arabidopsis thaliana gb|P25855.
 It contains a glycine cleavage H-protein domain
 PP|01597. ESTs gb|R90208, gb|AI
 Identities = 30/91 (32%), Positives = 53/91 (57%), Gaps = 1/91 (1%)

Query: 18 TISLTPELQDDLSTGVGYEFTD-DANLEVDVILNIRASKTVMAILSPILGKVVKNVTA 75
 TI +T QD LG V +VE +++++ + +E+ K ILSP++G+V++VNT
 Sbjct: 59 TIGITDHAQHLGELVVVFVELPEANSSVSKEKSPQAVSEVKATSEILSPISGEVIEVNTYL 118

Query: 77 SQEPTLLNSEKADENMLVVLTEVDYAAFEAL 107
 ++ P L+NS ++ W++ + A EAL
 Sbjct: 119 TESPLINSSPYDGMKRVKPSFARLEAL 149

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6183> which encodes the amino acid
 sequence <SEQ ID 6184>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3544(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2248-

An alignment of the GAS and GBS proteins is shown below.

Identities = 80/110 (72%), Positives = 98/110 (88%)

```

5  Query: 1  MKKIANLYLLIERNEELYTISLTPLEQDDLGTVGYVEFTDDANLEVDVILNIRASKTVMA 60
      MKKIANLYLLIEK ++ YTIS+TPELQDD+GT+GY EPTD+ +L VDD+ILN+PASKTVM+
      Sbjct: 1  MKKIANLYLLIEKTDKRYTISMTPELQDDIGTIGYAEFTDNDHLAVDDIILNLEASKTVMS 60

10  Query: 61  ILSPITGKVKVNTAASQSEPTLLNSEKADENLWLVLTVEDYAAFEALENA 110
      +LSPL G VV+ N AA+ PTLNSEKA+ENW+VVLV+VD AAP+ALE+A
      Sbjct: 61  VLSPLAGAVVERNEAATLTPTLNSEKASRNWIVLTVDDQAAFDALDA 110

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 2001

A DNA sequence (GBSx2111) was identified in *S. agalactiae* <SEQ ID 6185> which encodes the amino acid sequence <SEQ ID 6186>. This protein is predicted to be LRP16 (b1045). Analysis of this protein sequence reveals the following:

```

20  Possible site: 17
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.0608 (Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
25           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

30  >GP:AAF15294 GB:AF202922 LRP16 [Homo sapiens]
    Identities = 73/171 (42%), Positives = 98/171 (56%), Gaps = 13/171 (7%)

    Query: 88  DICLQVDAIVNAANSKLLQCPINHHICIDNQIHTFSGSRIRLACHOLMTQQGRMEAVQC 147
      DI L+VDAIVNAANS LLG +D IH AG L C L++ G+
      Sbjct: 78  DITKLEVDAIVNAANSLLQ-----GGVVDGCIHRAAGPLITDECKTLQSCCK-----TGK 127

35  Query: 148  AKLTSTSYHLPCYVIHTVGPYVKVDQKPSRIRIEDLLKSSYKSCQLAVRANLKTIVFPCT 207
      AK+T Y LP KYVIHTVGP + S+ E L+S Y S L L + L+++ FPCI
      Sbjct: 128  AKITOGYRLPAKYVIHTVGPYAYGEPASASQAAE--LRSCYLSLDDLEHRLRSVAFPCI 185

40  Query: 208  STGEGFPNQFAELAVQAILRWORENQHKL-VIIPNTPTPKDQDIYQKLL 257
      STG FG+P + AA+ + + EW +++ K+ +I F KD+DIY+ L
      Sbjct: 186  STGVPGYPCFAAARIVLATIRBKLQHKDKVDRLIICVPLEKDEDIVRSRL 236

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6187> which encodes the amino acid sequence <SEQ ID 6188>. Analysis of this protein sequence reveals the following:

```

45  Possible site: 25
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.1992 (Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
50           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

55  Identities = 139/266 (52%), Positives = 178/266 (66%), Gaps = 6/266 (2%)

    Query: 1  MPNQKQLLLAMIEYLOSEKLTVDL---RTIDLTQVWRGLWQDQPNIGQYLSLSD 56
      MP+ LL MI LQ+E+LT T Q +WR L+NQ+ +S++YL+LSD

```

-2249-

Sbjct: 1 MPSSFLLSEMIGLLQTSQLTSSWACPLNALTKRQDLWRALINQRPALPLSKYLNLE 60

Query: 57 RYLSHWNTQKVKITDVCHQTVYSNVFTYHGDICLLQVDAIVNAANSKLGCFINHEICT 116
YL W + ++ C +T Y++F YHGD L VDAIVNAANS+LLGCF PNH CI

5 Sbjct: 61 AYLDNHRASFPVSVKIDQKTNYSPLFLYHGDITRYLAVDAIVNAANSELGCFSPNHEICT 120

Query: 117 DNQHTFFAGSRRLRLAQHMTQQSRMRAVGQAKLTSYHLPCYVHTVGPYVKVDQKPS 176
DN HTFFAGSEIRLAC +MT+QSR EA+GQAKLT +YHLP Y+HTVGP + S

10 Sbjct: 121 DNAHTFFAGSRRLRLAQHMTQQSRKKAIGQAKLTSYHLPASYSITHTVGPRTKGHHVS 180

Query: 177 RIRKDLLKSSYSKCLQLAVRANLKITVFPCTSYGERGFPNQRAELAVQAILEWQRENQH 236
IR DLL Y+S L LAV+A L ++ F ISTGEGFP + AA++A++ +L+WQ E+

15 Sbjct: 181 PTRADLLARCYSRLSLDLAVKAGLTSIAFCSISTGEGFPFKKAAQIAIKTVLWQAEHP 240

Query: 237 K--LYIIFNTFTPKDQDIYQKLLKE 260
L IFNTFT +D+ +Y L KE

20 Sbjct: 241 SKYLTITFTNTFTSKDXYLYTYQKE 266

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2002

A DNA sequence (GBSx2112) was identified in *S.galactiae* <SEQ ID 6189> which encodes the amino acid sequence <SEQ ID 6190>. Analysis of this protein sequence reveals the following:

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2171 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
30 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6191> which encodes the amino acid sequence <SEQ ID 6192>. Analysis of this protein sequence reveals the following:

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2477 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
40 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 218/284 (76%), Positives = 250/284 (87%)

45 Query: 4 WKTLKKNHNSQSKRLSQLIBESDAIVVGIGAGMSAADGFTYIGPKRFRAFPFOITAKYQL 63
W T + N +Q+E L+QLI+E+DA+VVGIGAGMSAADGFTYIG RPE AFPOITAKYQ L
Sbjct: 4 WTTYFQKNLTQAEQLQLIKADALVVGIGAGMSAADGFTYIGSRFETAFFDFIAKYQFL 63

50 Query: 64 DMLQASLYDFEDNERYWAFQSRFVALNYLDQFVGQAYLDLKLAKKEYHIITNADNAF 123
DMLQASLI+DFEDN+KYWAFQSRFVALNYLDQFVGQ+YLDLKL+L K+YHIITNADNAF
Sbjct: 64 DMLQASLIQDFEDNORYWAFQSRFVALNYLDQFVGQSYLDLKLKELGTGDIYHIITNADNAF 123

55 Query: 124 AVADYNLEKVPFHIGQEGYGLWQCSQHCQQTYYRNDQAIRQMLAQQKDNKIPSNLIPKCPKC 183
VA Y+ +FHIGQEGYGLWQCSQHCQQT+D IRQMLA+QK+HK+P LIP CP+C
Sbjct: 124 WVAGYDPHNI FHIGQEGYGLWQCSQHCQQTYYKUDTVIRQMLAQKDNKIPSGQLIPKCPK 183

Query: 184 DQPFINKRNRKRGVEDADFAHQRYENFLSQHNDKVLNLEIGVGHITTPQFIKHPFW 243
+ DPFINKRNRKRGVEDADFAHQ+ RYE FL+H+ KVLNLEIGVGHITTPQFIKHPFW

-2250-

Sbjct: 184 EAPPEINKRMEKGMVEDADPHAQKARYAFLSEHKEGKVLYLEIGVGHTTPOFIKHPFW 243

Query: 244 RFLVSLIENSFLVTLNHHKYRIPQKIRSRVQLTQHARLIARAK 287
+ VS N N+LFLVTLNHHKYRIP 1R +S++L+HIA+LI+ K

Sbjct: 244 KRVSBNPRLFLVTLNHHKYRIPLSIRRQSLRLETHIAQLISATK 287

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2003

- 10 A DNA sequence (GBSx2113) was identified in *S. agalactiae* <SEQ ID 6193> which encodes the amino acid sequence <SEQ ID 6194>. Analysis of this protein sequence reveals the following:

Possible site: 21
>>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1086(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB12865 GB:Z99109 similar to lipote-protein ligase [Bacillus subtilis]
Identities = 130/331 (39%), Positives = 206/331 (61%), Gaps = 5/331 (1%)

- 25 Query: 9 NGKRITDGAIALNQVYILQNVLDDDDILFFYYCDPKVEIGKFNQAVIETNQYLKESDI 68
+ + I D I LA+ Y ++++ + L Y P + IGK QN + E N +Y++E+ I
Sbjct: 5 DNQININDPRINIAIEEYCVKHLDPQQYLLFVYNQPSIILQKQNTIREINFKYVEENG 64

- Query: 69 FVVRDITGGGAVYDGSAGNICYLNKDHQG-FGDFFKAYEPAIKALKITLGASSVEMRERN 127
VVRR +GGGAVY D G +N ++ KD G F +FK+ EP I+AL LG + E+ RN
30 Sbjct: 65 IIVVRLSGGGAVYHDLGNLNFSPITKDDGDSFNFKKFTRPVIQALHQLGVEA-ELSGRN 123

- Query: 128 DLVIDGKKVSGAANTIVNGRIYGGYSLLLDVDFDAMEKVLNPNRKIESKGKESVSRVSG 187
D+V+DG+K+SG A GRI+ +L+ D D + L + KIESKGKES+RSKV
35 Sbjct: 124 DIVVDGRKISGNAQFATKGRIFSHOTLMPDGAIDHVVSALKYKDKIESKGKESIRSRVA 183

- Query: 188 DIRSHLSRDYRHITDOPKCLMVCQLLHIDHIDQAKRYHLTEKDWAIADALADEKYGWD 247
+I L + +T++F+ ++ + + + + Y LTKDW I ++ R+Y+NWD
Sbjct: 184 NISEFLDK---MTTRESHLRLRHIFNTNDVGNVPEYKLTEKDWETTHQISKERYQND 240

- 40 Query: 248 WNYGNSPQYSYHRDARFPGQTYDPHLEIEKGIITNCRIYQDFFSSKDISDIENLLIGCPM 307
WNYG SP+++ + R+P G+ D HLE++KG I +C+I+GDFF D+S+IENLL+G
Sbjct: 241 WNYGRSPKPNLNFHSKRYPVGSDILHLEVKKGKIEDCKIFGDFVGVDVSEIENLLVGKQY 300

- Query: 308 KEELVLEKLSLTSLEDYFGQTSPEIKAVLP 338
+ ++ + L ++L+ YPG + E+ ++
45 Sbjct: 301 ERGVIADVLEGVNLKHYFGNITKEDFLDIY 331

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6195> which encodes the amino acid sequence <SEQ ID 6196>. Analysis of this protein sequence reveals the following:

- 50 Possible site: 21
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
bacterial cytoplasm --- Certainty=0.0939(Affirmative) < succ>
55 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 248/339 (73%), Positives = 283/339 (83%)

-2251-

Query: 1 MYLIEPIRNGKRITDGAIALAMQVYILQNVFLDDDLFPYYCDPKVEIGKFCNAVETIQ 60
 MYLIEPIRNGKRITDGAIALAMQVY+ +N+FLDDDLFPYYCDPKVEIGKFCNAV+ETIQ
 5 Sbjct: 1 MYLIEPIRNGKRITDGAIALAMQVYVQENLFLDDDLFPYYCDPKVEIGKFCNAVETIQ 60

Query: 61 EYLKEHDIPVVRRTGGGAVTVDSGAVNICYLMDKHGQFGDFKRAYEPAIKALKTIGASS 120
 EYLKEH IPVVRRTGGGAVTVDSGAVNICYL+ D+G FGDPR Y+PAI+AL IGA+
 Sbjct: 61 EYLKEHHPVVRRTGGGAVTVDSGAVNICYLINDNGIFGDFKRTYQDAIEALHLEGATE 120

Query: 121 VEMSRNDLVIDGKRVSGAAMTIVNGRIYGGYSLLLDVDFDAMEKVLNPKRKIESKGIK 180
 VEM RNDLVIDGKRVSGAAMTI NGR+YGGYSLLLDVDF+AMEK L PNRKIESKGI+
 10 Sbjct: 121 VEMSGRNDLVIDGKRVSGAAMTITANGRVYGGYSLLLDVDFEAMEKALPNRKIESKGIK 180

Query: 181 SVRSRVGDIRSHLSBDYRHITTDQPKLMVQQLLHIDHIDQAKRYHLTERKWAALDALAD 240
 SVRSRVG+IR HL+ Y+ IT ++FKDLMVQQLL I+ I QAKRY LTERKW IDAL+
 15 Sbjct: 181 SVRSRVGNIREHLAFTQGTTIEFKDLMVQQLLQISTISQAKRYDLTERKWAQIDALTE 240

Query: 241 EKYKNMDWNYGNPQYSYHRDARFPGSTYDPLHLEKGIITNCRIYDGFPPSKDISDIEN 300
 KY NW+VNYGN+PQY YHRD RF GT D HL+I+KG I CRIYDGF DI+++E
 20 Sbjct: 241 RKYHNEVNYGNAPQYRYHRDGRFTGTVDIHLDIKGYIAACRIYDGFPGKADIAELG 300

Query: 301 LLIGCPMKBELVLEKLTSLSDYFGQTSPEELKAVLFS 339
 LIG M++E VL L+ + L Y G + EE+ ++FS
 25 Sbjct: 301 HLIQTRGKEDVLATINADIAAPYLGAITAELEGDLFS 339

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2004

A DNA sequence (GBSx2114) was identified in *S. agalactiae* <SEQ ID 6197> which encodes the amino
 30 acid sequence <SEQ ID 6198>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.49 Transmembrane 196 - 212 (196 - 212)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.1595(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAE49329 GB:U39612 formyl-tetrahydrofolate synthetase
 [Streptococcus mutans]
 Identities = 432/556 (77%), Positives = 493/556 (87%)

45 Query: 1 NKTDIRIAQSVALKPIATVQVGIGFDDIELYGYKAKLSFDKIEAVKSGQVKQLLVT 60
 NKTDIRIAQSV L+PI +V+++GI FDD+ELYGYKAKL+FDKI+AV+ GKIL+LVT
 Sbjct: 1 NKTDIRIAQSVDLRPTTWWKKLIGDFDLELYGYKAKLTFDKIKAVEBNAPGKQLLVT 60

Query: 61 AINPTPAGEBKSTMSIGLADALNKGKKTMTALREPSLGPMVHGKGGAGGGYQVQLPME 120
 AINPTPAGEBKST++IGLADALNKGKKTMTA+REPSLGPMVHGKGGAGGGYQVQLPME
 50 Sbjct: 61 AINPTPAGEBKSTTTIGLADALNKGKKTMTALREPSLGPMVHGKGGAGGGYQVQLPME 120

Query: 123 DINEHPTGDMFAITTANNALSALLDNHIGQNELIDQQRVIVWKRVDINDRALRQVIVG 180
 DINEHPTGDMFAITTANNALSAL+DNH+HQGNEL IDQQR+IWKRVVDINDRALR V VG
 55 Sbjct: 123 DINEHPTGDMFAITTANNALSALINDHHLHQGNELIDQQRITWKRVDINDRALRQVIVG 180

Query: 181 LGSPVNGIPREDGFDITVASEIMAILCLATLSDLKRLSNIVVAYSRRNKPIYVCLKI 240
 LGSP+NGIPREDGFDITVASEIMAILCLAT+ DLK+RL+NTIV+ Y +R P+TV+DL++
 60 Sbjct: 181 LGSPINGIPREDGFDITVASEIMAILCLATNVDELKRLANTVIGRVDSPTVVDLLEV 240

Query: 241 BQALTLILKOTIKPNLVQTIYGTPLVHGGPFANTARGGNSVLATSTALRLADTVVTEAG 300
 +GAL LILK+ IKPNLVQTIYGTPL VHGPPFANTARGGNSVLATSTALRLADY +TEAG

-2252-

Sbjct: 241 QGALALILKEAIKPHLVQTIYGTFAFVHGGPFANIAHGQNSVLATSTALRLADYTTITEAG 300
 Query: 301 FGADLGAEKFLDIKTWNLPSTDAIVIVATLRALRMHGGVSKEDLSQENVRBAVKRGTNL 360
 FGADLGAEKFLDIK FNLPSTDAIVIVATLRALRM+GGV+K+ L+QENVRBAVK GF NL
 Sbjct: 301 FGADLGAEKFLDIKAPNLPSTDAIVIVATLRALRMHGGVAKDALNQENVRBAVKAGFANL 360
 Query: 361 ERHVNMRQYGVPPVVAINQFTADTESIATLTKLCSNIDVAELASVWEDGAGGLELA 420
 RHV NMR+YGVPPVVAIN+P DT EIA L+ LC+ IIV VELA+VW +GADGG+LA
 Sbjct: 361 ARHVENMRQYGVPPVVAINPFTDTNDEIAVLRLCAAIDVPVLEASVWANGADGGVDLA 420
 Query: 421 QTVANVIETOSSNYERLYNDEDTIEKIKKIVTKIYGGNKVHFGPKAQIKHKSNDNGWD 480
 T+ N IE S+YERLY++ ++SEK+ +I +Y +KV F KA+ Q+ + NGWD
 Sbjct: 421 NTLINTIRNPSHYRLLYDNRLSVBERVTEIAKELIYRADKVIPEKKAKTQIAQIVKNGWD 480
 Query: 481 KMPICMAKTCYSFSDNPGLGAPTDFTDITVREFVFKTGAGFIVALTGDLVIMPGLPKPKPA 540
 +FICMAKTCYSFSD+P LLGAPT FDT+RE VPK GAGFIVALTGSDV+TMPGLPKPKPA
 Sbjct: 481 NLPICMAKTCYSFSDDPKLLGAPTGFDTITRELVPKLGAGFIVALTGSDVIMPGLPKPKPA 540
 Query: 541 ALNMDVLEDGTAIGLF 556
 ALNMDV DGT+GLF
 Sbjct: 541 ALNMDVAADGTALGLF 556

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6199> which encodes the amino acid sequence <SEQ ID 6200>. Analysis of this protein sequence reveals the following:

25 Possible site: 50
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.49 Transmembrane 196 - 212 (196 - 212)
 ----- Final Results -----
 30 bacterial membrane --- Certainty=0.1595 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

35 >GP:AA849329 GB:U39612 formyl-tetrahydrofolate synthetase
 [Streptococcus mutans]
 Identities = 432/556 (77%), Positives = 490/556 (87%)
 Query: 1 MKSDIEIAQSVLQPIITDIVEKVGIDGDDIELYKGYKAKLSPEKMKAVANEPEKLIIVT 60
 MK+DIEIAQSV L+PIT++VKK+GID DD+ELYGKIKAKL+F+K+KAVE N PGKL+LVT
 Sbjct: 1 MKTDIEIAQSVLRPIITNVKRLGIDFDDLELYKGYKAKLTFDKIKAVENAPGKLIIVT 60
 Query: 61 AINPTPAGBKSTSTIGLADALNMGKKTWLAIREPLGPMGKIGQAGGGTAQVLPM 120
 AINPTPAGBKST++IGLADALN++GKKTW+A+REPLGPMGKIGQAGGGTAQVLPM 120
 Sbjct: 61 AINPTPAGBKSTITIGLADALNKGKKTWLAIREPLGPMGKIGQAGGGTAQVLPM 120
 Query: 121 DINLHPTGDMHAITTANNALGALIDNHLQQGNELGIDPRRIIWKRVLDNRLRQVTV 180
 DINLHPTGDMHAITTANNALGALIDNHL QSN+LGID RRIIWKRV+DNLRCAL V VG
 Sbjct: 121 DINLHPTGDMHAITTANNALGALIDNHLQQGNELGIDQRRRIIWKRVLDNRLRQVTV 180
 Query: 181 LGSPPVNGVPREDGFDITVASEIMAILCLATLDKLDKRLRIDVAVYITDRKPVVVDLKV 240
 LGS+NG+PREDGFDITVASEIMAILCLAT++DLK+RLA+IV+ Y +DR PTVVDL+V
 Sbjct: 181 LGSPPINGIPREDGFDITVASEIMAILCLATNVEDLKERLANIVIGYRFRSPVTVVDLKV 240
 Query: 241 EGRALILKDAIKPHLVQTIYGTFAFVHGGPFANIAHGQNSVLATSTALRLADYTTITEAG 300
 +GRL LILK+AIKPHLVQTIYGTFA+HGGPFANIAHGQNSVLATSTALRLADYTTITEAG
 Sbjct: 241 QGALALILKEAIKPHLVQTIYGTFAFVHGGPFANIAHGQNSVLATSTALRLADYTTITEAG 300
 Query: 301 FGADLGAEKFLDIKAPNLPSTDAIVIVATLRALRMHGGVAKDALNQENVRBAVKAGFANL 360
 FGADLGAEKFLDIK FNLP+PDA+VIVATLRALRM+GGVAK L EN EAV+ GFANL
 Sbjct: 301 FGADLGAEKFLDIKAPNLPSTDAIVIVATLRALRMHGGVAKDALNQENVRBAVKAGFANL 360
 Query: 361 ERHVNMRQYGVPPVVAINPFTDTNDEIAVLRLCAAIDVPVLEASVWANGADGGVDLA 420
 RHVENMR++ VPPVVAINP+ DT EIA L+ LC I VPELASVWANGAGGGGLA
 Sbjct: 361 ARHVENMRQYGVPPVVAINPFTDTNDEIAVLRLCAAIDVPVLEASVWANGADGGVDLA 420

-2253-

Query: 421 KTVVRVIDQEAADYKRLYSDBTLREKVINIVTQIYGGKAVQGPQAKTQLKQPAEFGMD 480
 T++ I+ + YKRLY + ++EEKV I +Y V F KAKTQ+ Q + GMD
 Sbjct: 421 NTLINTIENNPNHYKRLYDNNLSVREKVEITAKETRYADKVIPEKCAKTOQLAQIVKNGMD 480

Query: 481 KLPVCAKTCQYSFSDNPGLGAPTFDFTIREFVFKTGAGFIVGLTGDVWIMPGLEPKKPA 540
 LP+CAKTCQYSFSD+P LLGAPT FDTIRE VPK GAGFIV LTGDVWIMPGLEPK PA
 Sbjct: 481 NLPICMAKTCQYSFSDDPKLLGAPTFDFTIRELVFKLGAGFIVALTGDVWIMPGLEPKKPA 540

Query: 541 AMAMDVAENGTAIGLF 556
 A+ MDVA +GTALGLF
 Sbjct: 541 ALMDVAADGTAIGLF 556

An alignment of the GAS and GBS proteins is shown below.

Identities = 452/556 (81%), Positives = 513/556 (91%)

Query: 1 MKTDIRIAQSVALKPTARIVEQVIGPDDIELGYKAKLSFDKISAVKSQKVGLLIAT 60
 MK-DIRIAQSVAL+PI +IV++VGI DDIELGYKAKLSF+K++AV++ + GKLIVAT
 Sbjct: 1 MKSDIRIAQSVALQPTDITVKKVIGDGDITELGYKAKLSFERMKAVEANEPGKLLIAT 60

Query: 61 AINPTPAGEBKSTMSIGLADAINKIGKKTMLAREPSLGPMGIKGGAAGGGYAQLPME 120
 AINPTPAGEBKSTMSIGLADAIN++GKKTMLAREPSLGPMGIKGGAAGGGYAQLPME
 Sbjct: 61 AINPTPAGEBKSTMSIGLADAINQMGKKTMLAREPSLGPMGIKGGAAGGGYAQLPME 120

Query: 121 DINLHFTGDMHAITANNALSALLDNHIIHQNELDIDQRFVIMKRVVLDINDRALRQVIVG 180
 DINLHFTGDMHAITANNALSAL+DHH+ QGN+L ID RR+IWKRV+DINDRALRQVIVG
 Sbjct: 121 DINLHFTGDMHAITANNALSALIDNHILQGGNDLGDIPRIIWKRVLDINDRALRQVIVG 180

Query: 161 LGSFVNGVPREDGFDITVASEMAILCLATDLGDLKKRLSNIVVAYSRRNKPIYVDLKI 240
 LGSFVNG+PREDGFDITVASEMAILCLATDL DLKRL++IVVAY+ +RKP+YV+DLK+
 Sbjct: 161 LGSFVNGVPREDGFDITVASEMAILCLATDLKDLKKRLADIVVAYTYDRKPYVVDLKV 240

Query: 241 BGALILKOTIKPHLVQTYGTALPVHGGPFANIAHGCNSVLATSTALRLADYVTEAG 300
 BGALILKID IKPHLVQTYGTALP+HGGPFANIAHGCNSVLATSTALRLADYVTEAG
 Sbjct: 241 BGALILKDAIKPHLVQTYGTALPHGGPFANIAHGCNSVLATSTALRLADYVTEAG 300

Query: 301 FGADLGAEKFLDIKTNLPTSPDAIVIVATLRALMHGGVSKEDLSQENVEAVKRGFTNL 360
 FGADLGAEKFL+IK FNLP +FDAIVIVATLRALMHGGV+K DL+ EN EAV+ GF NL
 Sbjct: 301 FGADLGAEKFLNITKPNLKPADIAIVIVATLRALMHGGVAKSDLAEN+EAVALGFANL 360

Query: 361 ERHVNMRRQYGVPPVVAINGQFADTSESIATLKLCNIDVAVELASVWEDGAGGLALA 420
 +RHV NMRQ+ VPPVVVAI++F ADTS+SIATLK LC I V VELASVW +GA+GGL LA
 Sbjct: 361 ERHVNMRRQYGVPPVVAI+NEFVADTSESIATLKALCEIKVPPVVELASVWAGAGGLALA 420

Query: 421 QTVANVETQSSNYKRLNDEPTIREKIKKIVIKYGGKAVHFGPKAQIOLKPEPSNDG 480
 +TV VI+ +++YKRLY+DEPT+EEK+ IVT+IYGG V FGPKA+ QLK+P++ GMD
 Sbjct: 421 KTVVRVIDQEAADYKRLYSDBTLREKVINIVTQIYGGKAVQGPQAKTQLKQPAEFGMD 480

Query: 481 KNPICMAKTCQYSFSDNPGLGAPTFDFTIREFVFKTGAGFIVGLTGDVWIMPGLEPKKPA 540
 K+P+CAKTCQYSFSDNP+LLGAPTFDFTIREFVFKTGAGFIV LTGDV+IMPGLEPK PA
 Sbjct: 481 KLPVCAKTCQYSFSDNPGLGAPTFDFTIREFVFKTGAGFIVGLTGDVWIMPGLEPKKPA 540

Query: 541 ALAMDVAENGTAIGLF 556
 A+ MDV E+CTA+GLF
 Sbjct: 541 AMAMDVAENGTAIGLF 556

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 9057> which encodes amino acid sequence <SEQ ID 9058>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.49 Transmembrane 516 - 532 (516 - 533)

----- Final Results -----
 bacterial membrane --- Certainty=0.1595 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-2254-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS sequences follows:

Score = 604 bits (1540), Expect = e-174
 Identities = 304/555 (54%), Positives = 389/555 (69%), Gaps = 2/555 (0%)

Query: 4 SDIERIANSVIMEPISKVADQLGIDKRALCLYGYKAKIDARQLVALKNKFDGKLLVTAI 63
 +DIEIA SV ++PI+++ -QGI + + LVGYKAK+ ++ A+K+ GKLLVTAI
 Sbjct: 3 TDIEIAQSVALKPFIARIVBQVGIFDDIELYGYKAKLSFDKIEAVSQVVKGLLVTAI 62

Query: 64 SPTPAGEGKITTSGVLVDALSAIGKKAVIALREPSLDOXXXXXXXXXXXXXXXXXPMEDI 123
 +PTPAGEGK+T S+GL DAL+ IGKK +IALREPSL PMEDI
 Sbjct: 63 NPTPAGEGKSTMSIGLADALNKIGKKTMIALREPSLGPVVGIGGAAGGYAQLVPMEDI 122

Query: 124 NLHFTGDFHAIGVANNLLAALIDNHHHNSLIGDSRRITWKRVD+NDRLRHIVDGLQ 183
 NLHFTGD HAI ANN L-LAL+DNHH GN L ID RR+ WKRVD+NDR LR ++ GL
 Sbjct: 123 NLHFTGDMHAITANNALALLDNHHHQNELDIDQRVWKRVD+NDRLALRQIVTGLG 182

Query: 184 GKVGIPREDGYDITVASEIMAILCLSENISDLKALEKIIIGYNYQGFVXXXXXXXXX 243
 VNGIPREDG+DITVASEIMAILCL+ ++SDLK RL I++ Y+ +P+
 Sbjct: 183 SPVNGIPREDGFDTVASEIMAILCLATLSDLKRLSNIVVAYSNRKFIYVDLKEIG 242

Query: 244 XXXXXXXXXIHENLVQTLHTPALIHGGPFANIAHGNSVLATKLALKYDYVATEAGFG 303
 I RNLVQT+ TPAL+HGGPFANIAHGNSVLAT AL+ DY VTEAGFG
 Sbjct: 243 ALTLILKQITKENLVQITGTALVHGGPFANIAHGNSVLATSTALRLADYVTEAGFG 302

Query: 304 ADLGAERFIDIKRMSGLRPAAVLVATIRALKMHGVPKADLATENQAVVDGLPNLDK 363
 ADLGAERF+DIK P A+V+VAT+RALKMHGK K DL+ ENV+V G NI++
 Sbjct: 303 ADLGAERFLDITRNLPTSPDAIVVATIRALKMHGVSKEIDLQSNVEAVKRGFTNLER 362

Query: 364 HLANIQDVYGLFVVVAIRKFPLOTDALQAVYDACDKRGVDVVISDVWANGAGGRELAS 423
 H+ N++ YG+PVVAIN+F DT++E+ + C V V ++ VW +G GG ELA+
 Sbjct: 363 HVNMRQ-YGVFVVVAIRKFTADTSEIATLKTLCNIDVAVELASVWEDGADGGLLEAQ 421

Query: 424 KVVTLAE-QDNQFRFVYEDDSIETKTKIVTKVYGGKGINLSSAAKRELADLERIGFN 482
 V + E Q + ++ +Y ++D+IE K+ KIVTK+YGG ++ A+ +L + G+
 Sbjct: 422 TVANVETQSSNYKRLYNDEDTIEEKIKKIVTKIYGNKVHFGPKAQIQLKREFSDNGDK 481

Query: 483 YPICMAKTQYSFSDAKKLGAPTDFTVTISMLKVSAGAGFIVALTGAIMTMGLPKVPAS 542
 PICMAKTQYSFSD+ LGAPTD+T+ GAGFIVALTG ++TMEGLPK PA+
 Sbjct: 482 MPICMAKTQYSFSDNPHLLGAPTDFTDITVREFVPKIGAGFIVALTGVLTMGLPKPKPA 541

Query: 543 ETIDIDEEGNTIGLF 557
 +D+ E+G GLF
 Sbjct: 542 LNMVLEDDGTIGLF 556

SEQ ID 6198 (GBS131) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 6; MW 64.8kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 35 (lane 4; MW 90kDa).

GBS131-GST was purified as shown in Figure 201, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2005

A DNA sequence (GBSx2115) was identified in *S. agalactiae* <SEQ ID 6201> which encodes the amino acid sequence <SEQ ID 6202>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -10.03 Transmembrane 34 - 50 (29 - 56)

-2255-

INTEGRAL	Likelihood = -7.70	Transmembrane	90 - 106 (84 - 110)
INTEGRAL	Likelihood = -1.97	Transmembrane	62 - 78 (62 - 78)
INTEGRAL	Likelihood = -0.69	Transmembrane	275 - 291 (275 - 291)

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.5012 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

10 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA8609 GB:M37842 unknown protein [Streptococcus mutans]
 Identities = 243/373 (65%), Positives = 302/373 (80%), Gaps = 1/373 (0%)

15 Query: 71 IGAVLYLVNSEMDALSRVTLWLIVLAPLLGAMPLMTYKFDWGYRGLKQRLTLLDESQI 130
 IG+VLYLVNS+MD LS +TWL+++ P+LG +FL+YTK DWGYR LK ++ +
 Sbjct: 2 IGSVLYLVNSQMDTSLIITWLLVLPFPLIGTLFLIYTKQDWGYRGLKSLIKIGTKQAIKP 61

20 Query: 131 YLEDDPETINQLKSSTSTITTHLVQYFSAKHNFFVYRNTDVTFLPTGEAFPRMKKEELLK 190
 Y + D L +LK S + TY+L QY ++ G FPVY+NT VT- P G++ P+MK++LLK
 Sbjct: 62 YPQYDQRLIYKLNESHARTYNLAQYHRS-GGFPVYKNTKVTYPNGQSKPSEMKQLLAK 120

25 Query: 191 AKKYIFLEFFIIDEGIMNGEILSLBQKVEGVEVRILYDGMTEITKLSFDYTKRLEKIG 250
 A+K+IFLE+PII EG+MNGEILSLBQK+EGVEVR++YDGM+E+ LSPDY KRLKLEKIG
 Sbjct: 121 AEKPIFLEYFIIAEGIMNGEILSLBQKVEGVEVRVYDGMLELSTLSFDYAKRLEKIG 180

30 Query: 251 IKAKAPSPISPPFIISTYYNRDRHKIVVIDGVGMGGVNLADEYINIHILPGHWKDSGIM 310
 IKAK FSP+PF+STYYNRDRHKI+VID V GG+NLADEYIN IE PG+WKD+ +M
 Sbjct: 181 IKAKVSPSPITPPFSTYYNRDRHKILVIDNKVAFNGGINLADEYINIERGFYWKQTAVM 240

35 Query: 311 LKGGKAVDSFLLFLQWMSITEKMLVAPYLGVDHDLVENEQVYVIFYDSDPLTDKVGENV 370
 L+G+ V SF L+FLQWMS T + APYL + + GYVIFY DSDPLD +KVGENV
 Sbjct: 241 LRGEGVASFTIMFLQWSTTNKDYFAPYLTQNFHEIVANGYVIFYSDSDPLDHKVGENV 300

40 Query: 371 YIDILNHAREYVYINTPYLLDSEHAIQFAERGVDVRIIMPIGDPKPIPTALAKTY 430
 YIDILN AR+YVYINTPYLLDSE+HRA+QFAERGVDV+IIMPIGDPK +P+ALAK Y+
 Sbjct: 301 YIDILNQARDYVYINTPYLLDSEHAIQFAERGVDVRIIMPIGDPKVPFALAKRYF 360

Query: 431 QALTKSGVKIYEY 443
 AL +GVKIYE+
 40 Sbjct: 361 PALLDAGVKIYEF 373

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6203> which encodes the amino acid sequence <SEQ ID 6204>. Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -8.86	Transmembrane	84 - 100 (81 - 104)
INTEGRAL	Likelihood = -8.33	Transmembrane	28 - 44 (23 - 49)
INTEGRAL	Likelihood = -6.74	Transmembrane	56 - 72 (53 - 74)

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.4545 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

55 The protein has homology with the following sequences in the databases:

>GP:AAA23240 GB:J02911 formyltetrahydrofolate synthetase (FTHFS)
 (ttg start codon) (EC 6.3.4.3) [Moorella thermoacetica]
 Identities = 350/557 (62%), Positives = 438/557 (77%), Gaps = 2/557 (0%)

60 Query: 2 VLSIDIEIANSVTMEPISKVADQLGIDKBAALCYGKYKAKIDARQLVALKNPKDGLILVT 61
 VSDIRIA + M+P+ ++A LGI ++ +LYGKYKAKI LK+KPDGKLILVT
 Sbjct: 4 VPSDIEIAQAARKPKPVMELARGLGIDQDRVLYGKYKAKISLDVYRRLKDKPDGKLILVT 63

Query: 62 AISPTPAGBGKTTTSVGLVDALSAIGKKAVIALRSPSLGPFVFGVKGAGGGRHQVVPME 121

-2256-

		AI+PTPAGSGKITTSSVGL DAL+ +GK+ ++ LREPSLOP FG+KGGAAGGG+AQVVPME	
	Sbjct:	64 AITPTPAGSGKITTSSVGLTDALRLGKRVVCLREPSLOPFGIKGAAGGGYAQVVPME	123
5	Query:	122 DINLHFTGDFHAIQVANNLLAALIDNHHHNSLGIDSRITWKRVDNMNDRLRHIVDG	181
		DINLHFTGD HA+ A+MLAA++DNE+ GN L ID R ITW+RV+DNR LR+IV G	
	Sbjct:	124 DINLHFTGDIHAVTYAHNLLAAMVNLQGGVNLINDPKTITWKRVIDLNDRLRHIVIG	183
	Query:	182 LGQKVGIPREDGYDITVASEIMAILCISENISIDLKARLEKILIGYNYQGEVPTAKDLKA	241
		L GK NG+ERE G+DI+VASE+MA LCL+ ++ DLK R +1+GY Y G+PVTA DL+A	
10	Sbjct:	184 LGGRANGVPKRTGFDISVASEMACLCIASDLMOLKERPSKIVGVYTYDQKPVTAGDLEA	243
	Query:	242 GGLAALLKDAIHENLVQTLRHPTALIHGGPPANIAHGNSVLATKIALKYGDYAVTEAG	301
		G+A A L+KDAI ENLVQTL+TEA IHGGPPANIAHGNS++ATK ALK DY VTEAG	
	Sbjct:	244 QGSMALLMKDAIKENLVQTLTENTPAFIHGGPPANIAHGNSIATKIALKADYVTEAG	303
15	Query:	302 FGADLGAEKFIDIKRMSGRLPAAVLVATIRALKMGGVPKADLATENVQAVVDGLPNL	361
		FGADLGAEKF D+KCR +G +P A V+VAT+RALKMGGVFK+DLATEN++A+ +G NL	
	Sbjct:	304 FGADLGAEKFYDVCKRYAGFKPDATIVATVRLKMGGVFKSLDATENLEALKEGFANL	363
	Query:	362 DKHLANIQDVYGLFVVAINKFPDLTDAELQAVYDADRGGVDVVISDVWANGSAGGREL	421
		+KH+ NI +G+P VVAIN PV DT+AEL +Y+ C K G +V +S+VWA GG GG EL	
	Sbjct:	364 EKHENI -GKPGVEAVVAINAFPTDTAEALNLLYELCAFAGAEVALSEWAKSGGGEL	422
20	Query:	422 AEKV -TLAQDQNFRTYBEDDSIETKLKIVTKVYGGKIMLSAAKRELADLERLGF	480
		A KV+ TL + + P + Y D SI+ K K+ KI T++YG G+H+ ++ A + + E LG+	
	Sbjct:	423 ARKVLQTLSESRPSFHVNLNLDLSIKDKIAKIATETIGADGVNTHADKAIQRYSGY	482
	Query:	481 GNYPICHAKTYQSPSDAKKLGAPDTFTVVISNLKVSAGGFIVALGAIMTMGFLPKP	540
		GN P+ HAKTYQSPSD KLG P +FT+T+ ++SAG IV +TGAIMTMGFLPK P	
30	Sbjct:	483 GNLVVMHAKTYQSPSDHMTKLGPRNFNTITVREVLGAGGRLVETGAIMTMGFLPKP	542
	Query:	541 ASETIDIDEGNITGLF 557	
		A+ IDID +G ITGLF	
	Sbjct:	543 AACNIDIDAGVITGLF 559	
35		IGS:M37842 unknown protein [Streptococcus mutans] (v... 517 e-145	
		>GP:AAA8609 GB:M37842 unknown protein [Streptococcus mutans]	
		Identities = 246/370 (66%), Positives = 303/370 (81%), Gaps = 1/370 (0%)	
40	Query:	68 VLYLVNSMDAISRMIMLILIMTAPLQSLFLIYTKLDNGYRGLQRIHNLVDLSAPYLS	127
		VLYLVNS MD +S +TWL+I+ P+LQ+LFLIYTK DWGR LK I PY	
	Sbjct:	5 VLYLVNSQMDTISITWLLVILFPPIGLTFLIYTKDWGYSRELKSLIKKSTQAIKPYFQ	64
	Query:	128 DDAILEVLKSDSTTTHVLQVLESGRGNPFIYNNRTVTFPTGETFDSLKGFLPAKK	187
		D IL LK+S + TY+L QYL RS G FP+Y NP+VITYP G++ P+ +K+QL A+K	
45	Sbjct:	65 YDCRILYKLKESHARTYNLAQYLHRS -GGPPVYNNKTVTFYPPNGSKPSEMKKQLKAEK	123
	Query:	188 YIFLEFPIAEGQMWGEILSILEKVKSGVEVRVLPDGMNELSTLSDYAKRLBQIGIKA	247
		+IFLE+PIAEG MWGEILSILE+KV ROVEVRV++DGN ELSTLS DYAKRLB+IGIKA	
50	Sbjct:	124 FIFLEFPIAEGLMWGEILSILEQKVQEGVEVRVMYDGMLELETISFDYAKRLKIGIKA	183
	Query:	248 KSLFPISEFISTYYNRDHRKIVVIDCRVSPGTTGGINLADEYINFEVERPHWKDAGLMLEG	307
		K F PI+PF+STYYNRDHRKI+VID +V+P GGINLADEYIN++ERPG+WKD +NLEG	
	Sbjct:	184 KVPSPITPFVSTYYNRDHRKILVIDNKNVAPNGGINLADEYINQIERFPGYKDTAVMLEG	243
55	Query:	308 EATDSPLILFLQMWSSITEKKLIDPYLSDHSLKPSDGVVLPFGDSEPLDTDKIGKGVND	367
		E SP ++FLQMWSS T K+ PYL+ + ++ +GVVIFD DSEPL +K+G+NVVID	
	Sbjct:	244 EGVASPTLMFLQMWSSITNKDYEFAPYLTONFHEIVAKSVIPIYSDSEPLDHRKGVENVID	303
60	Query:	368 ILNHAKYEVYIMTPYLIIDSEMEHALQFAERGVVDIRIIMPGVDKGVYPALAKTYPAL	427
		ILN A++YVYIMTPYLIIDSEMEHAL+FA+ERGVDD+IIMPG+PK VP+ALAK Y+ AL	
	Sbjct:	304 ILNQARDVYIMTPYLIIDSEMEHALQFAERGVVDKIMPGIDPKKVPFALAKRYFPAL	363
	Query:	428 MSSGVKIYEX 377	
		+ +GVKIYE+	
65	Sbjct:	364 LDAGVKIYEF 373	

An alignment of the GAS and GBS proteins is shown below.

-2257-

Identities = 362/524 (69%), Positives = 437/524 (83%)

Query: 8 LISNKKVIVRLNKKSLRGIPSRRTTIVIAILLIQLQLFLASYSWLBQXRVMLATVEH 67
+I K K+ L+K K LRGIPSRRTT+I +L+ILQL+FL SY+W+BQXRVW+ +E
5 Sbjct: 2 IIRKKAQKYLHLHGKHGFLRGIPSRRTTIVLILLIQLQLFLASYSWLBQXRVMITLES 61

Query: 68 ILTIGAVLYLVNEMDALSRTVWLILVMIAPLIGAMFIYTKFDWGYRGLKQRLTLD 127
+ I VLYLVNS+MDA+SR+TWLIL+MIAPLLG++FL+YTK DWGYRGLKQ+ L+D
10 Sbjct: 62 VFAITTVLYLVNSMDAISRTVWLILMTIAPLLGSLFLIYTKLDWGYRGLKQRIHLVDL 121

Query: 128 SQIYLEDDEPETLNQLKSSTSTTYHLVQYFEKARGNFPVYRNIDVTFLPTGEAPFERMKKE 187
S YL DD L LK STSTTYHLVQY E++ GNFP+Y NT VI+ PTGE FF+ +KE+
15 Sbjct: 122 SAPYLSDDDAILEVLKSDSTSTTYHLVQYTLERSRGNFPYINNTVYTFPTGETFFDLSKQ 181

Query: 188 LLKAKKYIFLEFFI IDEGIMWGEILSILBQKVBEQVEVRILYDGMIEITKLSFDYTKRLE 247
L AKKYIFLEFFII EG MWGEILSILE+KV EGVEVR+L+DGM E++ LS DY KRLE
20 Sbjct: 182 LFLAKKYIFLEFFI IARGQMWGEILSILRKKVSRGVEVRVLFDMNLSLSSDYAKRLE 241

Query: 248 KIGIKAKAFSPISPFISTTYNYRDKRIVVIDGVMTGGVNLAD EYINHLFGHWKDS 307
+IGIKAK+F PISPFIISTTYNYRDKRIVVIDG V TGG+NLAD EYIN +E FGHWD+
25 Sbjct: 242 QIGIKAKSFLPISPFISTTYNYRDKRIVVIDGEVSTGGINLAD EYIN EYINVERFGHWKDA 301

Query: 308 GIMLKGKAVDSFLLFLQWMSITEKMLVAPYLGVDHDLVNEGYVPIPTGDSPLDTKVG 367
G+ML+G+A D SFL+LFLQWMSITE+++++ PYL H + ++GYVPIPTGDSPLDTDK+G
30 Sbjct: 302 GMLGEGATDSFLLFLQWMSITEKELI IDPYSLHSLKLPDSGVIPTGDSPLDTKIG 361

Query: 368 ENVYIDILNHARETVYINTPYLILDSRLSHAQFAAERGVDVRIIMPIDPKPIPYALAK 427
+NVYIDILNRA+ETVYIMTFYLLDSE+EHA++FA+ERGVD+RIIMP+G+PDK +PYALAK
35 Sbjct: 362 KNVYIDILNHARETVYINTPYLILDSBMEHALRFASERGVDIRIIMPVGPDGVPYALAK 421

Query: 428 TTYQALTKSGVKIYETLGFVHSKIFLSDNTKAVVGTINLDYRSLYHFECAVLYKVDA 487
TY+AL SGVKIYET GFVHSK+P+SDNTKAVVGTINLDYRSLYHFECA YLY+V
40 Sbjct: 422 TTYKALMSGVKIYETQGFVHSKVFISDNTKAVVGTINLDYRSLYHFECAVLYRVSV 481

Query: 488 IQDIYRDYMDTLNKSRLVSLKDIINNIPKQKIGIVIGITKFIAPLL 531
I DI D+ + +S L++ + P +QK+IG++ + IAPLL
45 Sbjct: 482 IADI VDPNBAQKQSLMSTSDHLTQRPWYQKLGILLVRIIAPLL 525

A related GBS gene <SEQ ID 8953> and protein <SEQ ID 8954> were also identified. Analysis of this protein sequence reveals the following:

LipoP Possible site: -1 Crend: 6
McG: Discrim Score: -8.80
GVH: Signal Score (-7.5): -1.94
Possible site: 53
45 >>> Seems to have no N-terminal signal sequence
ALOM program count: 4 value: -10.03 threshold: 0.0
INTEGRAL Likelihood = -10.03 Transmembrane 34 - 50 (29 - 56)
INTEGRAL Likelihood = -7.70 Transmembrane 90 - 106 (84 - 110)
INTEGRAL Likelihood = -1.97 Transmembrane 62 - 78 (62 - 78)
50 PERIPHERAL Likelihood = 1.22 199
modified ALOM score: 2.51
*** Reasoning Step: 3
55 ----- Final Results -----
bacterial membrane --- Certainty=0.5012(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

60 The protein has homology with the following sequences in the databases:

32.5/57.2% over 498aa

Bacillus firmus

SP|066043| CARDIOLIPIN SYNTHETASE (BC 2.7.8.-) (CARDIOLIPIN SYNTHASE) (CL SYNTHASE).
Insert characterized

65 GP|2952028|gb|AAC05444.1||U88888 cardiolipin synthase Insert characterized

ORF01572(409 - 1893 of 2193)
 SP|G66043|CLB_EACPI(5 - 503 of 503) CARDIOLIPIN SYNTHETASE (EC 2.7.8.-) (CARDIOLIPIN
 SYNTHASE) (CL SYNTHASE). GP|2952028|gb|AAC05444.1|U88888 cardiolipin synthase [Bacillus
 firmus]
 %Match = 17.9
 %Identity = 32.5 %Similarity = 57.1
 Matches = 162 Mismatches = 204 Conservative Sub.s = 123

153 183 213 243 273 303 333 363
 NLQLSIWMF**KTVOPLDYFK**RGRACDASLFLIGIRF*LEII**NNRMLEF*OYALIK*LIWRGEKLISNKVKIVRLNK

```

393      423      447      477      507      528      558      588
SKKSLLRGIPIRRTTVIAILLIQLLF--LIASYSMLQYRVMLATVEHILT--IGAVLYLVNSEMDALSrvTWLIIWMI
           : | : | || |   : | : | : |   : | | : : |   : | | : : |
MKNRLNVLAFFALLPAALYSIRGFLSNMVGLTSVVFITLSVFIFIGIILFFEN--RHPTKLTLLVLAA
          10       20       30       40       50       60

```

[illegible]

849 879 909 939 969 999 1029 1059
LPTGEAFFEKKELLKAKYIFLEPFIIDEGIMMGEILSILEQKVEEGVEVRILYDGMETITKLSPDYTKRLKEIKIGIKA
| : : : | : | : || : | : || : | : || | : || | : || | : || | : || :
LTDGKETVAHLQALMAHHHLEUYFIRVDDLGNQDKDILLSKAGEGVHVFPLYDG-VGSWLKSSEYVELRZAGVEM

160 170 180 190 200 210 220

1086 1116 1146 1176 1206 1236 1266 1293
KAFSPIS-PFISTYNNYRDHRIKVIVDVGVMGTGCVNLAEVDYIHHILFGHMKGIMLKGKAVDSFLLLQFNWMSI-TE
|::||:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
VSFPSPVKLPFLTHTITNHNRRHKIKVIDGVGVGFVGLGNLEVDYLKGDAFYGWDRTHLYVRGEAAVTQLQLIFLDWGYQTG

 240 250 260 270 280 290 300

1323 1353 1383 1413 1443 1473 1503 1533
EKMLVAPYLGVHDDLVENEGVYIPYGDSPLDTKVGENVYIIDLNHAREVYIMTPYLILDSLELHAIPAAREGVDVRI
| | | :
ETILNQYLSPSLSTIKGGDGQVMIAISGPDTRVEKNKLFFSMSTIAKSIIWASLPFYDDDLILSAKIALSAGIDVRI

320 330 340 350 360 370 380

[illegible]

1803 1833 1863 1893 1923 1953 1983 2013

YRDYDITLNRSLRVLSKDIIMNPKFKQKIVGIVTKIAPL*K*FIFNLILKVN*RI*LYLXKSGCILTKLC*TVVMR*VD

||:| | :||:| | :||:| | :||:| |

VSDYVYDLHSHNQINFSLFKNRPFPHRLIBSTRSLHSLPL

480 490 500

Q ID 8954 (GBS277d) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 150 (lane 18; MW 51kDa), in Figure 151 (lane 17 & 18; MW 51kDa) and in Figure 182 (lane 12; MW 51kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 151 (lane 15 & 16; MW 76kDa) and in Figure 58 (lane 5; MW 87kDa).

GBS277d-His was purified as shown in Figure 235, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2259-

Example 2006

A DNA sequence (GBSx2116) was identified in *S.agalactiae* <SEQ ID 6205> which encodes the amino acid sequence <SEQ ID 6206>. This protein is predicted to be aspartate-semialdehyde dehydrogenase. Analysis of this protein sequence reveals the following:

```

5      Possible site: 42
      >>> Seems to have an uncleavable N-term signal seq

      ----- Final Results -----
10      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9831> which encodes amino acid sequence <SEQ ID 9832>
was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA26850 GB:J02667 aspartate beta-semialdehyde dehydrogenase (EC
      1.2.1.11) [Streptococcus mutans]
      Identities = 261/357 (73%), Positives = 304/357 (85%), Gaps = 1/357 (0%)

20 Query: 1 MGYTVAIVGATGAVGTQMIRQLQESNLPBQVKLLSSSSRSAGKILHFKDEAIRVSETTKE 60
      MGYTVAIVGATGAVGT+MI-QLEQS LP++V+LLSSRSAGK+L +KD+ + VS TTK+
      Sbjct: 1 MGYTVAIVGATGAVGTRMIQQLQESLTPVDKVRLLSSRSAGKVLQYKDDQVTVSLTKD 60

25 Query: 61 SFYDVIDALFSAGGSISAKFAPYAVKSGAVVVDNTSYFRQNPDPVPLVPEVNAHAMHGN 120
      SF VDLALFSAGGS+SAKFAPYAVK+GAVVVDNTS+FRQNPDPVPLVPEVNA+AM HN
      Sbjct: 61 SFEAVDLALFSAGGSVSAKFAPYAVKAGAVVVDNTSHFRQNPDPVPLVPEVNAAYMDAHN 120

Query: 121 GIICPNCSTIQMIALEPIRQKNGIERVISTYQAVSGSGARAVESTKQELKQVLNENL 180
      GIICPNCSTIQM+ALEPIRQKNG+ RVISTYQAVSG+G A+ ET ++++v+ND +
30 Sbjct: 121 GIICPNCSTIQMVALEPIRQKNGLSRVISTYQAVSGAGQSAINSTVREIKVENVNIGV 180

Query: 181 SEDQLIATVLPSSDDKHYPIAFNALPQIDIFTDNDTYEEMKMTLSTKKIMEDATIKVS 240
      P + A + P D+KHYPIAFNAL QID+FTDNDTYEEMKMT STKKIME+ + VS
35 Sbjct: 181 DPKAVHADIFPSGGDDKHYPIAFNALAQIINFDTNDTYEEMKMTNETKKIMEPELPVS 240

Query: 241 ATCVRIPIVLGSHSESIYIETKELASISEIKGAIANFPFQAVLQDLPSSQIYPCAINAVGHR 300
      A CVR+P+L HSE++YIETK++A I E+K AIA FPGAVL+D QIYPQA NAVG R
40 Sbjct: 241 AHCVRVPIPLFSSHSEAVYIETKIDVAPISEVKAAIAAFPGAVLEDDIKHQIYPCQANAVGSR 300

Query: 301 ETPVGRIRKDLQENGVMHVVSDNLLKGAWNSVQIAETLHKGKGLVPAKELKPEL 357
      TFPVGRIRKDL ENG+HMMVVSNDLLKGAWNS+ A LH+ GLV+ ELKPEL
45 Sbjct: 301 TFPVGRIRKDLDIENGHMMVVSNDLLKGAWNSIITANRLHGRGLVSTSELKPEL 356

```

No corresponding DNA sequence was identified in *S.pyogenes*.

45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2007

A DNA sequence (GBSx2117) was identified in *S.agalactiae* <SEQ ID 6207> which encodes the amino acid sequence <SEQ ID 6208>. Analysis of this protein sequence reveals the following:

```

50      Possible site: 22
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL Likelihood = -2.66 Transmembrane 33 - 49 ( 33 - 49)

      ----- Final Results -----
55      bacterial membrane --- Certainty=0.2062(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

-2260-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 500.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2008

A DNA sequence (GBSx2119) was identified in *S.agalactiae* <SEQ ID 6209> which encodes the amino acid sequence <SEQ ID 6210>. Analysis of this protein sequence reveals the following:

10 Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.3853 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2009

A DNA sequence (GBSx2120) was identified in *S.agalactiae* <SEQ ID 6211> which encodes the amino acid sequence <SEQ ID 6212>. This protein is predicted to be unnamed protein product (clpP). Analysis of
25 this protein sequence reveals the following:

Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.3863 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- A related GBS nucleic acid sequence <SEQ ID 10061> which encodes amino acid sequence <SEQ ID
35 10062> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6213> which encodes the amino acid sequence <SEQ ID 6214>. Analysis of this protein sequence reveals the following:

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.2682 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

45

An alignment of the GAS and GBS proteins is shown below.

Identities = 175/196 (89%), Positives = 187/196 (95%)

-2261-

Query: 5 MI PVVIEQTSRGRSYDIYSRLKKDRIIMLTGQVEDNMANSITAOQLFLDAQENTKDIYL 64
 Sbjct: 1 MI PVVIEQTSRGRSYDIYSRLKKDRIIMLTGQVEDNMANSITAOQLFLDAQENTKDIYL 60

5 Query: 65 YVNTFGGSVAGLAIVDTMNFIKADVQTTVMGMAASMTT+IASSG KGRFMLPNAEYMI 124
 YVNTFGGSVAGLAIVDTMNFIKADVQTTVMGMAASMTT+IASSG KGRFMLPNAEYMI 120

10 Query: 125 HQPMGGTGGGTQQ+DMAIAAEHLKTRHLEKILADNSGQSEIKVHDADERDRWMSAQET 184
 HQPMGGTGGGTQQ+DMAIAAEHLKTRHLEKILADNSGQSEIKVHDADERDRWMSAQET 180

Query: 185 LDYGFIDAIMENNLQ 200
 L YGFID IMEN L+

15 Sbjct: 181 LAYGFIDEIMENNLK 196

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2010

- 20 A DNA sequence (GBSx2121) was identified in *S. agalactiae* <SEQ ID 6215> which encodes the amino acid sequence <SEQ ID 6216>. This protein is predicted to be uracil phosphoribosyltransferase (upp). Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

25 INTEGRAL Likelihood = -0.43 Transmembrane 127 - 143 (127 - 144)
 INTEGRAL Likelihood = -0.06 Transmembrane 72 - 88 (72 - 89)
 INTEGRAL Likelihood = -0.06 Transmembrane 154 - 170 (154 - 170)

----- Final Results -----

30 bacterial membrane --- Certainty=0.1171 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- A related GBS nucleic acid sequence <SEQ ID 10063> which encodes amino acid sequence <SEQ ID 10064> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA26890 GB:L07793 uracil phosphoribosyltransferase
 (Streptococcus salivarius)
 Identities = 192/209 (91%), Positives = 202/209 (95%)

40 Query: 1 MKGFQVISHPLIQHKLILRRRTTSTKDFRELVDRIAMMGYEVSERDLPLEVEIQTPVA 60
 MKGFQVISHPLIQHKLILRR TSTKDFRELVAETAMMGYEVSERDLPLE+VEIQTP+
 Sbjct: 1 MKGFQVISHPLIQHKLILRRRTTSTKDFRELVDRIAMMGYEVSERDLPLEVEIQTPIT 60

45 Query: 61 TTVQKQLAGKKLAIVPILRAGIMVDGFLSLVPAKVGHGMVREDESTPQVEYLKLE 120
 TVQKQLAGKKLAIVPILRAGIMVDGFLSLVPAKVGHGMVREDEST +PVEYLKLE
 Sbjct: 61 KTVQKQLAGKKLAIVPILRAGIMVDGFLSLVPAKVGHGMVREDESTLPVEYLKLE 120

50 Query: 121 DIDQRQIFVVDPMATGGSAILAVDSLKKRGAASIKFVCLVAAPBGVAALQEMHDPDIY 180
 DIDQRQIFVVDPMATGGSAILAVDSLKKRGA+IKFVCLVAAPBGV LQ+MHP+DIY
 Sbjct: 121 DIDQRQIFVVDPMATGGSAILAVDSLKKRGAANIKFVCLVAAPBGVKLQDMHDPDIY 180

Query: 181 TAALDEKLINHGVIYVPLGDAGDRLPGTK 209
 TA+LDEKLIN+GYIVPGLGDAGDRLPGTK

55 Sbjct: 181 TASLDEKLINHGVIYVPLGDAGDRLPGTK 209

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6217> which encodes the amino acid sequence <SEQ ID 6218>. Analysis of this protein sequence reveals the following:

-2262-

Possible site: 26
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.59 Transmembrane 72 - 88 (72 - 89)
 INTEGRAL Likelihood = -0.22 Transmembrane 127 - 143 (127 - 144)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein is similar to uracil phosphoribosyltransferase from *S.salivarius*:

>GP:AAA26890 GB:L07753 uracil phosphoribosyltransferase [Streptococcus salivarius]
 Identities = 191/209 (91%), Positives = 205/209 (97%)

Query: 1 MGKQVISHPLIQHKLILRRQTTSKDFRELNVNKLAMLAGYEVSRDLPLEDDVDTQTPVS 60
 MGK QVISHPLIQHKLILRR TSTKDFRELNVNKLAMLAGYEVSRDLPLE+V+IQTP++
 Sbjct: 1 MGKQVISHPLIQHKLILRRDSTKDFRELNVNKLAMLAGYEVSRDLPLEVEIQTTPIT 60
 Query: 61 KTVQKQLAGKKLAIVPILRAGIMVDGLLSLVPAKVGHIGMYRNEETLEPVEYLKLP 120
 KTVQKQL+GKKLAIVPILRAGIMVDG LSLVPAKVGHIGMYR+EETLEPVEYLKLP
 Sbjct: 61 KTVQKQLSGKKLAIVPILRAGIMVDGFLSLVPAKVGHIGMYRDEETLEPVEYLKLP 120
 Query: 121 DIHQRIQIFLVDPMILATGGSAILAVDSLKKRGAANIKFVCLVAAPBGVKQLQEAHPDIDIF 160
 DI+QRQIF+VDPMLATGGSAILAVDSLKKRGAANIKFVCLVAAPBGVKQL+AHDPIDI+
 Sbjct: 121 DIDQRQIFVDPMLATGGSAILAVDSLKKRGAANIKFVCLVAAPBGVKQLQEAHPDIDIY 180
 Query: 181 TAALDHLNENHYIVPGLGDAGDRLPGTK 209
 TA+LD+ LNE+GYIVPGLGDAGDRLPGTK
 Sbjct: 181 TASLDEKLNENHYIVPGLGDAGDRLPGTK 209

An alignment of the GAS and GBS proteins is shown below.

Identities = 190/209 (90%), Positives = 201/209 (95%)
 Query: 1 MGKQVISHPLIQHKLILRRQTTSKDFRELNVNKLAMLAGYEVSRDLPLEDDVDTQTPVA 60
 MGK QVISHPLIQHKLILRR TSTKDFRELNV+EIAMLAGYEVSRDLPLEDDV+IQTPV+
 Sbjct: 1 MGKQVISHPLIQHKLILRRQTTSKDFRELNVNKLAMLAGYEVSRDLPLEDDVDTQTPVS 60
 Query: 61 TVQKQLAGKKLAIVPILRAGIMVDGFLSLVPAKVGHIGMYRDEETPQVEYLKLP 120
 TVQKQLAGKKLAIVPILRAGIMVDG LSLVPAKVGHIGMYR+EET +PVEYLKLP
 Sbjct: 61 KTVQKQLAGKKLAIVPILRAGIMVDGLLSLVPAKVGHIGMYRNEETLEPVEYLKLP 120
 Query: 121 DIDQRQIFVDPMLATGGSAILAVDSLKKRGAANIKFVCLVAAPBGVAALQEAHPDIDY 160
 DI+QRQIF+VDPMLATGGSAILAVDSLKKRGAANIKFVCLVAAPBGV LQEAHPD+DI+
 Sbjct: 121 DIDQRQIFLVDPMILATGGSAILAVDSLKKRGAANIKFVCLVAAPBGVKQLQEAHPDIDIF 180
 Query: 181 TAALDHLNENHYIVPGLGDAGDRLPGTK 209
 TAALD+ LNEHGYIVPGLGDAGDRLPGTK
 Sbjct: 181 TAALDHLNENHYIVPGLGDAGDRLPGTK 209

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2011

A DNA sequence (GBSx2122) was identified in *S.agalactiae* <SEQ ID 6219> which encodes the amino acid sequence <SEQ ID 6220>. This protein is predicted to be hemolysin (patB). Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.29 Transmembrane 88 - 104 (86 - 106)

-2263-

bacterial membrane --- Certainty=0.2317 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB15133 GB: Z99120 aminotransferase [Bacillus subtilis]

Identities = 130/381 (34%), Positives = 221/381 (57%), Gaps = 4/381 (1%)

10 Query: 5 DFTSLPERFSSNTTKKAVQK---DQELPLWADMDFFIFPMSEATDFSHQWVGVD 61
 +F ER + ++KW + + LP+W+ADMDF ++EA+++ +FGY
 Sbjct: 2 NFDKRRERLGTQGVKMDKTGELGVTDAIPMWVADMDFAPEATRALKKRLDHGIGYT 61

15 Query: 62 SPKDSLYQAISNWEVQEHGYQFPRKSLILLIDGVVPAISVAIQAFTRKGDVALLINTPVYFP 121
 +P A+ W HG+ + +S+ GVV A+S+A+QAF+ GD V+ + PVY P
 Sbjct: 62 TPDQCKTCAVGVWQNRHGWKVPESITFSGVVITLGNVAVQAFTEPGDQVVQVPVYTF 121

20 Query: 122 FARTIKYNNRHVLSNGLNNQYFIDPFQLEKDIIENNVKLYIFCSPHNPGGRVVTGGE 181
 F ++ N RE++ N LL + + IDF+ LE + + +V L+I C+PHN GR W++ +
 Sbjct: 122 FYHVEKNGREILHNPLEKQAYADPEDLETKLSDPSVTLFILCNPHNPGGRWSRED 181

25 Query: 182 IQKIGDICKRYNVILVSDIEHQDLVLFDFNVHSFNTVDSSFKELSVLSSATKTFNIAGT 241
 + K+G++C + V +VSDIEH DL+L+ + H F ++ F ++SV ++ +KTFNIAG
 Sbjct: 182 LLKLGELCLEHGVTVVSDIEHSDLMYGHKHTFPASLSDDFADISVTCAPSKTFNIAGL 241

30 Query: 242 KNSFAIIENEKLRSDFKKQIANNQBEISSLGALLATEVAFTRKQKMLKALMELESSIEY 301
 + S II + R+ F N +++ + A E A++K WL L +E ++
 Sbjct: 242 QASAIIFPRLKRAKFSASLRNGLGGLNAFAVTAIEAYSKGFGLDELITYIEIQDNE 301

35 Query: 302 LYEQL-TQKTEIKVMKPGSTYLWLDPSAYNLTHLEIQKRLYDAKLLINDGLTFQKRGK 360
 L T+ +K+MKP+ +YL+WLDPSAY L+ E+Q+++ K+IL G +G G+
 Sbjct: 302 AEAFLSTELPKVIMKPDASYLWLDPSAYGLSDAELQRMKKKGVILEPGTKYPGGE 361

Query: 361 KHARINVAAPRSVIEAVLRL 381
 R+N + +++ + R+
 Sbjct: 362 GPMRLNAGCSLTLQDGIRRI 382

There is also homology to SEQ ID 1006.

SEQ ID 6220 (GBS392) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 2; MW 46.4kDa). It was also expressed in *E. coli* as a GST-fusion
 40 product. SDS-PAGE analysis of total cell extract is shown in Figure 83 (lane 5; MW 71kDa).

GBS392-GST was purified as shown in Figure 217, lane 4.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2012

45 A DNA sequence (GBSx2123) was identified in *S. agalactiae* <SEQ ID 6221> which encodes the amino acid sequence <SEQ ID 6222>. This protein is predicted to be tRNA methylase, SpoU family (cspR). Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1436 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

55

The protein has homology with the following sequences in the GENPEPT database.

-2264-

>GP:AB02738 GB:U58864 CspR [Bacillus subtilis]
Identities = 84/155 (54%), Positives = 120/155 (77%), Gaps = 3/155 (1%)

5 Query: 19 H I V L F E P Q I P A N T G N I A R T C A A T N A P L H I I R P M G F F I D D K M K R A G L D Y N D K L D V S F Y D G 78
H-VL++P+I PANTGN IARTCAATN Lh+IRP4GF DDK +KRAGLDYH++V++D
Sbjct: 4 H V V L Y Q P E I P A N T G N I A R T C A A T N I T L I L R P L G F S T D D M L K R A G L D Y W E F V N V Y H D S 63

10 Query: 79 L E E - F M L S C R G K V H L I S K F A D K V S D E N Y N D - D Q D H Y F M F G R E D K G L P E T F M R E H A E K A L 136
LEE F +GK I+KF +++ Y D D+D+F+FGRE GLP+ +++ ++ L
Sbjct: 64 L E H L F E A Y K G K P F F I T K F Q Q Q P H T S F D Y T D L D E Y F V F V G R E T S G L P A D L I Q A N M D R C L 123

15 Query: 137 R I P M N D E H V R S L N V S N T V C M I V Y E A L R Q Q S F P N L E 171
R+PM EHVRSIN+SNT ++VYEALRQQ++ +L+
Sbjct: 124 R L P M T - E H V R S I N L S N T A I L V Y E A L R Q Q N Y R D L K 157

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6223> which encodes the amino acid sequence <SEQ ID 6224>. Analysis of this protein sequence reveals the following:

Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2236(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 135/182 (74%), Positives = 150/182 (82%)

30 Query: 1' M N I E T L T Q K N H R S D S G R N H I V L F E P Q I P A N T G N I A R T C A A T N A P L H I I R P M G F F I D D K M 60
M + L K N + R N H I V L F + P Q I P N T G N I A R T C A A T N A P L H I I + P M G F F I D D + K M
Sbjct: 13 M T K E L I N K N D K V K K A R N H I V L F Q P Q I P Q N T G N I A R T C A A T N A P L H I I K P M G F F I D D K M 72

35 Query: 61 K R A G L D Y N D K L D V S F Y D G L E E P M L S C R G K V H L I S K F A D K V S D E N Y N D Q D H Y F M F G R E D 120
K R A G L D Y N D K L + F Y D L E + F + C G + H L I S K F A Y S Y D H Y F + F G R E D
Sbjct: 73 K R A G L D Y N D K L E H F Y D H L E Q F I N Q C H S Q L H I S K F A V N N Y S Q A T Y A D G D S H Y F L F G R E D 132

40 Query: 121 K G L P E T F M R E H A E K A L R I P M N D E H V R S L N V S N T V C M I V Y E A L R Q Q S F P N L E L S H T Y E N D K 180
G L P E F M R E H A E K A L R I P M N D E H V R S L N V S N T V C M + Y E A L R Q Q F L E L H T Y R + D K
Sbjct: 133 T G L P E D F M R E H A E K A L R I P M N D E H V R S L N V S N T V C M I V Y E A L R Q Q G P Q L E L H T Y E N D K 192

45 Query: 181 L K 182
L K
Sbjct: 193 L K 194

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2013

A DNA sequence (GBSx2124) was identified in *S. agalactiae* <SEQ ID 6225> which encodes the amino acid sequence <SEQ ID 6226>. Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -6.79	Transmembrane	82 - 98 (69 - 100)
INTEGRAL	Likelihood = -6.46	Transmembrane	27 - 43 (24 - 47)
INTEGRAL	Likelihood = -5.52	Transmembrane	132 - 148 (126 - 151)
INTEGRAL	Likelihood = -5.10	Transmembrane	162 - 178 (161 - 185)

----- Final Results -----

bacterial membrane --- Certainty=0.3718(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9411> which encodes amino acid sequence <SEQ ID 9412> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:CAB13143 GB:Z99110 similar to amino acid permease [Bacillus subtilis]
Identities = 46/143 (32%), Positives = 81/143 (56%), Gaps = 1/143 (0%)

Query: 3 FAYDQWTFVNIAPEVKNPCKNIPLAPVIGPALILLSYLAFFYGLTQILGASPIMTTND 62
FAYDGN + + E+KRP+K LP A G ++ Y+ + L IL A+ I+T G +
10 Sbjct: 203 FAYDWILLALGEMKNPEKLLPRAMTGGLLIVTAIYIFINPALLHLISANEIVTLGEN 262

Query: 63 AINFYANNIIIFGSPVGRLLSFIVILSVLGVANGELLGTMRLPQAPAEKGVK-SERMANTIN 121
A + AA ++FG G+L+S +I+S+ G NG +L R+ A AER + +E+++++
15 Sbjct: 263 ATSTAATMLFGSIGGKLISVGIIVSIFGCLANGKVLSPFRVS FAMAERKOLPFAEKLSHVH 322

Query: 122 LKYQMSLPASLTVTVAIFWLFV 144
++ A A+A+ ++
Sbjct: 323 PSFRTFWIAISFQIALALIMMLI 345

20 There is also homology to SEQ ID 3114.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2014

25 A DNA sequence (GBSx2125) was identified in *S. agalactiae* <SEQ ID 6227> which encodes the amino acid sequence <SEQ ID 6228>. Analysis of this protein sequence reveals the following:

Possible site: 20
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----

bacterial cytoplasm ---	Certainty=0.1849 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

35 A related GBS nucleic acid sequence <SEQ ID 9439> which encodes amino acid sequence <SEQ ID 9440> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD23454 GB:AF117741 cochaperonin GroES [Streptococcus pneumoniae]
Identities = 31/52 (59%), Positives = 42/52 (80%)

40 Query: 2 GDGIRITLIGELVAPSVABSDTVLVENGAGLEVKGDKNEKVTTVRSEDIVAVVK 53
G G+RTL G+LVAPSV GD VLVE AGL+VKDG+EK +V E++I+A++
Sbjct: 42 GQSVRTLNGDLVAPSVTGDKVLVKAHAGLVDKDKDKYIIVGKNTLAIIE 93

45 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6229> which encodes the amino acid sequence <SEQ ID 6230>. Analysis of this protein sequence reveals the following:

Possible site: 50
>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----

bacterial cytoplasm ---	Certainty=0.3290 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

-2266-

An alignment of the GAS and GBS proteins is shown below.

Identities = 29/49 (59%), Positives = 39/49 (79%)

```

5 Query: 4 GIRTITGELVAPSVAGDITVLVNGAGLEVKGNGKVTTVRESIDIVAVV 52
      G+RT+TG+ V PSV+ G VLVNG LEV +EKV++RESDI+A+V
  Sbjct: 60 GVRTITGDSVLPSVVGQVLVNGKHDLVTVDDKVSLLRESIDIAIV 108

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 2015

A DNA sequence (GBSx2126) was identified in *S. agalactiae* <SEQ ID 6231> which encodes the amino acid sequence <SEQ ID 6232>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

      bacterial cytoplasm --- Certainty=0.1272 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD23455 GB:AF117741 chaperonin GroEL [Streptococcus pneumoniae]
Identities = 472/539 (87%), Positives = 513/539 (94%), Gaps = 1/539 (0%)

```

25 Query: 1 MAKDIKFSADARSAMVRGVLDILADTVKVLGPKGRNVVLEKAFSGPLITNDGVTIAKEIS 60
      M-K+IKFS+DARSAMVRGVLDILADTVKVLGPK RNVVLEK+FGSPLITNDGVTIAKEIS
  Sbjct: 1 MSKRIKFPSSDARSAMVRGVLDILADTVKVLGPKDRNVVLEKSPGSGPLITNDGVTIAKEIS 60

30 Query: 61 LEDHPFNMGAKLVSEVASKINDIAGDGTDTATVLTQAIIVREGLKQNVAGANPIGIRRGIE 120
      LEDHPFNMGAKLVSE+ASKINDIAGDGTDTATVLTQAIIVREG+KNVTAGANPIGIRRGIE
  Sbjct: 61 LEDHPFNMGAKLVSE IASKINDIAGDGTDTATVLTQAIIVREGLKQNVAGANPIGIRRGIE 120

35 Query: 121 TAVSAAVEELKKEIAQPVSGKEAIAQVAVSSRSEKVGGEYSRAMERVMNDGVTITIEESRG 180
      TAV+AAVE LK A PV+ KEAI+QVAVSSRSEKVGGEYSRAME+VG DGVITIEESRG
  Sbjct: 121 TAV+AAVEALKNNAI PVANKEAISQVAVSSRSEKVGGEYSRAMEKVGQGVITIEESRG 180

40 Query: 181 METELEVEVGMQFDRGYLSQYMTDNEKMWSELENFYILITDKKISNIQILPLEELVK 240
      METELEVEVGMQFDRGYLSQYMTVD+EKW++LENFYILITDKKISNIQILPLEE +L+
  Sbjct: 181 METELEVEVGMQFDRGYLSQYMTVDSEKMWADLENFYILITDKKISNIQILPLEELIQ 240

45 Query: 241 TNRPLLI IADVDGEALPTVLINKIRGTFNVVAVKAPGPGDRRKAMLEDAIILTGTVVT 300
      +NRPLLI IADVDGEALPTVLINKIRGTFNVVAVKAPGPGDRRKAMLEDAIILTGTV+T
  Sbjct: 241 SNRPLLI IADVDGEALPTVLINKIRGTFNVVAVKAPGPGDRRKAMLEDAIILTGTVIT 300

50 Query: 301 EDLGLDLKDATMQVLQSAKVTVDKDSTVIVEBGAGDSSAIANKVAIIKSQMSATSDPDR 360
      EDLGL+LKDAT+ LQ+ A+VTVDKDSTVIVEBGR+ AI+RVA+IKSQ+E TTS+PDR
  Sbjct: 301 EDLGLGLKDATTEALQSAKVTVTVDKDSTVIVEBGRNPEAISHRVAVIKSQIETTSPDR 360

55 Query: 361 EKLQRLAKLAGGVAVIKVGAATTELKEMKLRIEDALNATRAAEVREGIVSGG3TALNV 420
      EKLQRLAKL+GGVAVIKVGAATTELKEMKLRIEDALNATRAAEVREGIV+GG3TAL NV
  Sbjct: 361 EKLQRLAKLGGGVAVIKVGAATTELKEMKLRIEDALNATRAAEVREGIVAGG3TALNV 420

Query: 421 TEKVAALKINGDESTRNIVLRALKEFPVRQIAYNAGYEGSSVIERLKQSEIGTGPMANG 480
      I A L+L GDE TGRNIVLRALKEFPVRQIA+NAG+EGS++I+RLK +E+G GPMAG G
  Sbjct: 421 TPAAETLTLGDEATGRNIVLRALKEFPVRQIAHNGAFEGSIVIDRLKNAELGIGPMAGT 480

Query: 481 EWNVMVTGGIIPVKVRSALQNAASVASLILITRAVAVANKPEBPAPAMPDPSMGG 539
      ENV+M+ GIIDPVK+RSALQNAASVASLILITRAVAVANKPE AP APAMPDPSMGG
  Sbjct: 481 EWNVMIDQGIIDPVKVSRSALQNAASVASLILITRAVAVANKPEPAP-APAMPDPSMGG 538

```

-2267-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6233> which encodes the amino acid sequence <SEQ ID 6234>. Analysis of this protein sequence reveals the following:

```

Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1070 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 491/543 (90%), Positives = 515/543 (94%), Gaps = 3/543 (0%)

Query: 1  MARDIKFSADARSAMVRGVLDIADTVKVTLPKGRNVVLEKAFGSLTINDGVITAKSIE 60
Sbjct: 3  MARDIKFSADAR+AMVRGVD+LADTVKVTLPKGRNVVLEKAFGSLTINDGVITAKSIE 62

Query: 61  LEDHFENMGAKLVEVASKINDIAGDGTITATVLTQAVI BGLKNVTAGANFIGIRGIE 120
Sbjct: 63  LEDHFENMGAKLVEVASKINDIAGDGTITATVLTQAVI BGLKNVTAGANFIGIRGIE 122

Query: 121  TAVSAAVEELKEIAQPVSGKEAIAQVAVSSRSEKVEYISEAMERVNDGVITIEESRG 180
Sbjct: 123  TATATAVEALKIAQPVSGKEAIAQVAVSSRSEKVEYISEAMERVNDGVITIEESRG 182

Query: 181  MTELEVVEGCMQFDRGYLSQYMTDNEKQV++LENP+ILITDKK+SNIQ+ILPLEEVK 240
Sbjct: 183  MTELEVVEGCMQFDRGYLSQYMTDNEKQVADLENP+ILITDKK+SNIQ+ILPLEEVK 242

Query: 241  TNRPLLIADDDVGEALPTLVINKIRGTNNVAVKAPGDRRKQMLEDIALITGGTVIT 300
Sbjct: 243  TNRPLLIADDDVGEALPTLVINKIRGTNNVAVKAPGDRRKQMLEDIALITGGTVIT 302

Query: 301  EDLGLDLKDATMQLQQAQKTVDKDSTVIVEGSGSSAIAANRVAIKSQMEATTSDFDR 360
Sbjct: 303  EDLGLDLKDATMQLQQAQKTVDKDSTVIVEGSGSSAIAANRVAIKSQMEATTSDFDR 362

Query: 361  EKLQERLAKLAGGVAVIKVGAATETELKEMKLRIEDALNATRAAVEGIVSGGTALVNV 420
Sbjct: 363  EKLQERLAKLAGGVAVIKVGAATETELKEMKLRIEDALNATRAAVEGIVSGGTALVNV 422

Query: 421  IEKVAALKINGDEETGRNIVLRALKEPVRQIAINAGYRGSV++LK S GTGPHAA G 480
Sbjct: 423  IEKVAALLEGEDATGRNIVLRALKEPVRQIAINAGYRGSVVVDIKQSPAGTGFNAATG 482

Query: 481  BVDVMTTGLIIPDKVTRTSALQNAASVASLILTTAVVANKPEP--EAPTAPA-MDPMM 537
Sbjct: 483  BVDVMTTGLIIPDKVTRTSALQNAASVASLILTTAVVANKPEP AP PA MDP MM 542

Query: 538  GGF 540
Sbjct: 543  GGF 545

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2016

A DNA sequence (GBSx2127) was identified in *S.agalactiae* <SEQ ID 6235> which encodes the amino acid sequence <SEQ ID 6236>. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have no N-terminal signal sequence

```

-2268-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3216 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10247> which encodes amino acid sequence <SEQ ID 10248> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06113 GB:AP001515 transcriptional regulator (GntR family)
 [Bacillus halodurans]
 Identities = 50/171 (29%), Positives = 86/171 (50%), Gaps = 17/171 (9%)

Query: 21 HVQVYNKIPNMIQDGTYSFGMLPSEPELAGQINVSATLRKSLALLQEDHLVKNIRKSG 80
 ++QV +K+ + + + G Y G +LPSE EL+ QL VSRATLR++L LL+E+ +V G G
 Sbjct: 10 YLQVIDKLKHDMEAGVYEEGEKLPSEFELSKQLGVSATLRKSLALLLEEGLVVRHVG 69

Query: 81 NFIRENSNLSSENGYENRQHPITCLTSKITEVELE-----PRVEVPASAITASLKQ 132
 F+ + + L G E + +T I + +E + + +E +
 Sbjct: 70 TFV--HTKPLFSAGIEELY-----SVTDMIRHADMEPGTIFLSSYQIEATDDKKRFQTD 122

Query: 133 ETPVVVIADRWYHTDDGFLAYTLSPFPIELISDASISLHDTKQLNFIIEG 183
 +++ +R D P+ Y L +P ELI + S+H+ +L+ +E G
 Sbjct: 123 NLQGLMIRVETADGVPIVYCLDKLAELI--GQSHVEHNSLDHLESG 171

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6237> which encodes the amino acid sequence <SEQ ID 6238>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2297 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 154/244 (63%), Positives = 189/244 (77%)

Query: 7 MPKNEIANKANKLKHVQVYNKIPNMIQDGTYSFGMLPSEPELAGQINVSATLRKSLAL 66
 M N+L KL LKHVQVYN IF +IQDGTYSFGMLPSEPELAGQINVSATLRKSLAL
 Sbjct: 1 MSTNDLYTKLKLKLVQVYNTIPQLIQDGTYSFGMLPSEPELAGQINVSATLRKSLAL 60

Query: 67 LQEDHLVKNIRKSGNPFIRENSNLSSENGYENRQHPITCLTSKITEVELEPRVEVPASAI 126
 LQEDHLVKNIRKSGNFI + G+E QHP I L+S IT+VELE+A+VP AI
 Sbjct: 61 LQEDHLVKNIRKSGNPFILKTPETKYHQFVYLQHPIVASLSDITKYVELEYRIVETVAI 120

Query: 127 TASLKQETFPVVVIADRWYHTDDGFLAYTLSPFPIELISDASISLHDTKQLNFIIEG 196
 TASLKQETFPVV+I DRWYH+ + +AV+LSPITR+IS I+L+ +L P+EE IY+
 Sbjct: 121 TASLKQETFPVVVIADRWYHQNKAIAVSLSPFPIEIVISKYAINLQGEPLLTLEEKIYE 180

Query: 187 EGISSHSGHGLGVATQGNFGATKTYLSDHQQLFIQETLFPQEKILMCGHGVPIHPEL 246
 G +SHS +GV +GN+ATKYTLG+ FILIQET+ + IL+ KHYVP +P+L
 Sbjct: 181 SGKASHSCNQTGYTKCYTYMCKTYLSESAFILIQETLYNGKDLVSTKHVPAFLFDL 240

Query: 247 SITE 250
 + S
 Sbjct: 241 KVQS 244

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2017

A DNA sequence (GBSx2128) was identified in *S.agalactiae* <SEQ ID 6239> which encodes the amino acid sequence <SEQ ID 6240>. This protein is predicted to be purine nucleoside phosphorylase (udp-1). Analysis of this protein sequence reveals the following:

```

5   Possible site: 47
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10  bacterial cytoplasm --- Certainty=0.3910 (Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:AA65977 GB:AE001270 uridine phosphorylase (udp) [Treponema
   pallidum]
   Identities = 145/246 (58%), Positives = 171/246 (68%)

Query: 11 QYHLQIRPGDVGRYVIMPGDPKRCACIAEHFDNALVLVADSREYVITYTGLNGEKVST 70
   +YH+ ++ D+G YVI+PGDP R KIA+HF + V +REYVITYTGL VSV ST
20  Sbjct: 10 EYHIGLWASDIGHYVILPGDPARSEKIAQHFSPHKVGHNREYVITYTGLCETPVSMST 69

Query: 71 GIGGSPASIAMIELKLCGADTFIRVGTCCGIDLDVKGQDIVIATGAIRMBGTSKEYAPIE 130
   GIGGSP+I +EEL GA TPIRVGT GG+ D+ G +VIATGAIR BGTSKEYAP+E
25  Sbjct: 70 GIGGSPSTAIGVEELIHLGAHTPIRVGTSGGNQPDILAGTVVIATGAIRPBGTSKEYAPVE 129

Query: 131 FPAVADLEVINALVNAAKKLGYSHTAGVQCKDAFYQGHEPERMPVSYELLNKEAWKRL 190
   FPAV D VT AL +AA+ + GVVQCKD FYQGH P MFV EL KW AW
30  Sbjct: 130 FPAVEDFTVTAALKHAAEDVQVRALGVVQCKDNFYQGHSPHTMPVHAEITQRHAWAIAC 189

Query: 191 GTKASEMESAAALFVAASHLGVROSGDFLVVGNQERNALGMNPNNAHDTEAAIQVAEALR 250
   T ASEMESAAALFV S VR G+ LV+GNQ R A G+++ HYTE AI+VAVERA++
35  Sbjct: 190 NTLASEMESAAALFVLGVSRRVRTGAVLLVGNQTRRAQGLLEDIQVHDTEAIRVAEAVK 249

Query: 251 TLIEND 256
   LI D
35  Sbjct: 250 LLITQD 255

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6241> which encodes the amino acid sequence <SEQ ID 6242>. Analysis of this protein sequence reveals the following:

```

40  Possible site: 47
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
45  bacterial cytoplasm --- Certainty=0.3910 (Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

50  Identities = 259/259 (100%), Positives = 259/259 (100%)

Query: 1 MQNSGSEVGLQYHLQIRPGDVGRYVIMPGDPKRCACIAEHFDNALVLVADSREYVITYTGL 60
   MQNSGSEVGLQYHLQIRPGDVGRYVIMPGDPKRCACIAEHFDNALVLVADSREYVITYTGL
Sbjct: 1 MQNSGSEVGLQYHLQIRPGDVGRYVIMPGDPKRCACIAEHFDNALVLVADSREYVITYTGL 60

55  Query: 61 NGEKVSVTSTGIGGSPASIAMIELKLCGADTFIRVGTCCGIDLDVKGQDIVIATGAIRMS 120
   NGEKVSVTSTGIGGSPASIAMIELKLCGADTFIRVGTCCGIDLDVKGQDIVIATGAIRMS
Sbjct: 61 NGEKVSVTSTGIGGSPASIAMIELKLCGADTFIRVGTCCGIDLDVKGQDIVIATGAIRMS 120

60  Query: 121 GTSKEYAPIEFPAVADLEVINALVNAAKKLGYSHTAGVQCKDAFYQGHEPERMPVSYEL 180
   GTSKEYAPIEFPAVADLEVINALVNAAKKLGYSHTAGVQCKDAFYQGHEPERMPVSYEL
Sbjct: 121 GTSKEYAPIEFPAVADLEVINALVNAAKKLGYSHTAGVQCKDAFYQGHEPERMPVSYEL 180

```

-2270-

Query: 181 LNKWEAWKRLGTTKASEMSAALFVAASHLGVROGSDFLVVGQERNALGMDNPMADHTEA 240
LNKWEAWKRLGTTKASEMSAALFVAASHLGVROGSDFLVVGQERNALGMDNPMADHTEA
Sbjct: 181 LNKWEAWKRLGTTKASEMSAALFVAASHLGVROGSDFLVVGQERNALGMDNPMADHTEA 240

Query: 241 AIQVAVEALRTLIENDKQ 259
AIQVAVEALRTLIENDKQ
Sbjct: 241 AIQVAVEALRTLIENDKQ 259

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2018

A DNA sequence (GBSx2129) was identified in *S. agalactiae* <SEQ ID 6243> which encodes the amino acid sequence <SEQ ID 6244>. This protein is predicted to be nucleoside transporter. Analysis of this protein sequence reveals the following:

Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -9.45 Transmembrane 35 - 51 (30 - 57)
INTEGRAL Likelihood = -9.29 Transmembrane 8 - 24 (1 - 28)
INTEGRAL Likelihood = -8.07 Transmembrane 388 - 404 (379 - 404)
INTEGRAL Likelihood = -7.27 Transmembrane 104 - 120 (100 - 127)
INTEGRAL Likelihood = -6.58 Transmembrane 259 - 275 (255 - 284)
INTEGRAL Likelihood = -4.35 Transmembrane 172 - 188 (171 - 190)
INTEGRAL Likelihood = -3.50 Transmembrane 200 - 216 (199 - 221)
INTEGRAL Likelihood = -2.18 Transmembrane 352 - 368 (352 - 371)
----- Final Results -----
bacterial membrane --- Certainty=0.4779 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10245> which encodes amino acid sequence <SEQ ID 10246> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA05165 GB:AP001512 nucleoside transporter [Bacillus halodurans]
Identities = 160/405 (39%), Positives = 256/405 (62%), Gaps = 8/405 (1%)
Query: 5 MQFTYSITGILLVLGIVVAISFNKSVSLGKALIVQFIALLVIRIPLGQVSVV 64
M ++ ++G+I+V I +A S NR++ I L +Q + A+I++P GQ ++ ++
Sbjct: 1 MNILGILLGIVVFLIAFPTNRRRAIKPRTILGGLAIQLFAIVLKIPIAGQALLES 60
Query: 65 TGVTVKINGQGLNPFVFGSLADGCKTGFPAIQTLGNIVFLSALVSLIYVGLGFV 124
V +I+ G++FVFG + G+ GP+FAI L ++F SAL+S+LY+GI+ FV+
Sbjct: 61 NVVNLNLSYANBGIDFVFGGFFBFGSGVGVPFAINVLVSVIIFSSALISILYVGLGQFVI 120
Query: 125 KWKIGKVGKIMKSEVESFVAVANMPLGCTDSPLVSKYLGRTDSRIMVVLVSGMGMS 184
K IG + ++ +S+ ES A AN+P+G++P+V YL +M+ SE+ V+ G+ S+
Sbjct: 121 KIIGALSLLGLTSAKESMSAANIPVQCTEAPLVVVKPYLPKMTQSLFAVMTGGLASVA 180
Query: 185 VSLIGGYIALGIPMEYLLIASTMVPFGSILIAKILLPQTEPVQKI -DKIMNKGNANV 243
S+L GY LG+P++YLL AS M ++AK++P+TE DD K+ + N+
Sbjct: 181 GSVLIGYSLLGVPLQYLLAASFVAAPGLIMAKNIMPETKTDQEDDFKLAARDSTNL 240
Query: 244 IDAIEGASTGQAMFASIGASLIAFVGLVSLDMMISGLG-----IRLEQLFSYVFP 296
IDA A GASTG + +I A L+AFV L++LIN +L +G + LE I YVFP
Sbjct: 241 IDAAAGASTGLAVLNIAAMLAFVALIALINGILGWIGGLPGASQLSGLLILGYVFP 300
Query: 297 PGFLMGPHKNILLGEGNLLGSKLILNEFVSFQGLDLIKSLDYRTALVATISLQGFANLS 356
P++G L G+ +G KL++NEFV++ I++L + +V + +LQGFAN S

-2271-

Sbjct: 301 LAFVIGIFWAEALQAGSYIQGLVNVNEFVAYLSFAPETENLSKAVMVISFALGCFANFS 360

Query: 357 SLGICVSGIAVLCPKRGSTLARIIVFRAMIGGIASVMSLSAFIVGIV 401

SLGI + G+ L F +R +ARL RA++ G S+LSA I G++

Sbjct: 361 SLGILLGGLGKLAPSRPDTARIGLRRAILAGTLASLSASAGML 405

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6245> which encodes the amino acid sequence <SEQ ID 6246>. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -9.45 Transmembrane 35 - 51 (30 - 57)
 INTEGRAL Likelihood = -9.29 Transmembrane 8 - 24 (1 - 28)
 INTEGRAL Likelihood = -8.07 Transmembrane 388 - 404 (379 - 404)
 INTEGRAL Likelihood = -7.27 Transmembrane 104 - 120 (100 - 127)
 INTEGRAL Likelihood = -6.58 Transmembrane 259 - 275 (255 - 284)
 INTEGRAL Likelihood = -4.35 Transmembrane 172 - 188 (171 - 190)
 INTEGRAL Likelihood = -3.50 Transmembrane 200 - 216 (199 - 221)
 INTEGRAL Likelihood = -2.18 Transmembrane 352 - 368 (352 - 371)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4779 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

25 The protein has homology with the following sequences in the databases:

>GP:EAB05165 GB:AP001512 nucleoside transporter [Bacillus halodurans]
 Identities = 160/405 (39%), Positives = 257/405 (62%), Gaps = 8/405 (1%)

Query: 5 MDTYSIIGILLVLGIVVAISFNKRSVSLSLIGALIVQFIATILVRIPLQQQIVSV 64

M ++ ++GI+V I +A S NR+++ I L +Q + A+I++IP GQ ++ ++

Sbjct: 1 MNILMGLLGIIVVFLIAPAFSTNRRATKPTITLGGLAIQLLFAITVLKIPAGQALLESIT 60

Query: 65 TGVTSVINCQAGLNVFVPSGLADSGAKTGFIFAIQTLGNIVLSALVSLLYVYGILGFPV 124

V ++I+ G++FVFG + G+ GF+FAI L ++F SAL+S+LYT+GI+ FV+

Sbjct: 61 NVVLNIISYANGHIDFVFGFGFEBGSGVGFVFAINVLVSIVIFPSALISILYTLGIMQFVI 120

Query: 125 KWIGKGVGKIMKSSSEVSFVAVANMPLQITDPSILVSKYLRMTDSEIMVVLVSGMGMS 184

K IG + ++ +S+ ES A AN+P+GQT++P++V YL +NT SE+ V+ G+ S++

Sbjct: 121 KIIGGALSWLLGTSAESMSAANIPVQGTREAPLVVVKPFLPMTQSEFAVMTGGLSAVA 180

Query: 185 VSLIGYSLGIPMEYLLIATMVPISGILLIAKILLPQTEPVQKI-DDIKMDKNGSNAV 243

S+L GY LG+P++YLL AS M ++AK++P+TE DD K+ + N+

Sbjct: 181 GSVLIGYSLGIVPLQLLAASFMAPAGLIMAKMIMPETEKTTDAEDDFKLAKDEESTNL 240

Query: 244 IDAIEAGSTGQAQMSFISGASLTAFQGLVSLINMMLSGLG-----IRLEQIPSVFAP 296

IDA A GASTG + +I A L+AFV L++LIN +L +G + LE I YVFAP

Sbjct: 241 IDAARNGASTGIMLVNTAAMLLAFVALIALINGLAWIGGLGASQSLSLILGIVFAP 300

Query: 297 FGFIMGFDHKNILLESNLGSKLILNEFVSFQQLGHLIKSLDYRTALVATISLQGFANLS 356

F++G L G+ +G KL++NEFV++ I++L + +V +LHGFSAN S

Sbjct: 301 LAFVIGIFWAEALQAGSYIQGLVNVNEFVAYLSFAPETENLSKAVMVISFALGCFANFS 360

Query: 357 SLGICVSGIAVLCPKRGSTLARIIVFRAMIGGIASVMSLSAFIVGIV 401

SLGI + G+ L F +R +ARL RA++ G S+LSA I G++

Sbjct: 361 SLGILLGGLGKLAPSRPDTARIGLRRAILAGTLASLSASAGML 405

An alignment of the GAS and GBS proteins is shown below.

Identities = 399/404 (98%), Positives = 401/404 (98%)

60 Query: 1 MEVIMQFIYSIIGILLVLGIVVAISFNKRSVSLSLIGALIVQFIATILVRIPLQQQV 60

+EVIMQFIYSIIGILLVLGIVVAISFNKRSVSLSLIGALIVQFIATILVRIPLQQQV

Sbjct: 1 LEVIMQFIYSIIGILLVLGIVVAISFNKRSVSLSLIGALIVQFIATILVRIPLQQQV 60

Query: 61 SVVSTGVTKVINCQAGLNVFVPSGLADSGAKTGFIFAIQTLGNIVLSALVSLLYVYGIL 120

			SVSTSGVT VINCQAGNIFVPGSLADSGAKTRFPAITQIGNIVFSLVSLVLYYGIL	
	Sbjct:	61	SVSTSGVTGVSINCQAGNIFVPGSLADSGAKTRFPAITQIGNIVFSLVSLVLYYGIL	120
5	Query:	121	GFVVKIGKGVGKIMKKSSEVESFVAVNMPLGQDTSPLVSKYLGKMTDSEIMVVLVSGM	
			GFVVKIGKGVGKIMKKSSEVESFVAVNMPLGQDTSPLVSKYLGKMTDSEIMVVLVSGM	
	Sbjct:	121	GFVVKIGKGVGKIMKKSSEVESFVAVNMPLGQDTSPLVSKYLGKMTDSEIMVVLVSGM	180
	Query:	181	GSMVSISGGYTAGLIMPEYLLIASTMVPISGLIAKILLPQTEPVQKIDDKMNGNN	240
10			GSMVSISGGYTAGLIMPEYLLIASTMVPISGLIAKILLPQTEPVQKIDDKMNGNN	
	Sbjct:	181	GSMVSISGGYTAGLIMPEYLLIASTMVPISGLIAKILLPQTEPVQKIDDKMNGNN	240
	Query:	241	ANVIDAIAEGASTQMAFSIGASLIAPVGLSLINMMSGLGIRLEQIPSYVFPAGPL	300
			ANVIDAIAEGASTQMAFSIGASLIAPVGLSLINMMSGLGIRLEQIPSYVFPAGPL	
	Sbjct:	241	ANVIDAIAEGASTQMAFSIGASLIAPVGLSLINMMSGLGIRLEQIPSYVFPAGPL	300
15				
	Query:	301	MGFDHKNILLENILGSLKILINEVSPVQQLGDLKSLDYRTALVATISLCGFANLSSGI	360
			MGFDHKNILLENILGSLKILINEVSPVQQLG DLKSLDYRTALVATISLCGFANLSSGI	
	Sbjct:	301	MGFDHKNILLENILGSLKILINEVSPVQQLGDLKSLDYRTALVATISLCGFANLSSGI	360
20				
	Query:	361	CVSGIAVLCPKIRGTARLVFRAMIGGIAVMSLSAFTVIGVITLF	404
			CVSGIAVLCPKIR TARI VFRAMIGGIAVMSLSAFTVIGVITLF	
	Sbjct:	361	CVSGIAVLCPKIRSTLARLVFRAMIGGIAVMSLSAFTVIGVITLF	404

A related GBS gene <SEQ ID 8955> and protein <SEQ ID 8956> were also identified. Analysis of this protein sequence reveals the following:

```

30  Ldpop: Possible site: -1    Crend: 1
    McQ: Discrim Score:      13.83
    GVH: Signal Score (-7.5): -2.63
        Possible site: 25
    >>> Seems to have an uncleavable N-term signal seq
ALOM program count: 8 value: -9.65 threshold: 0.0
    INTEGRAL Likelihood = -9.45    Transmembrane 35 - 51 ( 30 - 57)
    INTEGRAL Likelihood = -9.29    Transmembrane 8 - 24 ( 1 - 28)
35  INTEGRAL Likelihood = -8.07    Transmembrane 388 - 404 ( 379 - 404)
    INTEGRAL Likelihood = -7.27    Transmembrane 104 - 120 ( 100 - 127)
    INTEGRAL Likelihood = -6.58    Transmembrane 259 - 275 ( 255 - 284)
    INTEGRAL Likelihood = -4.35    Transmembrane 172 - 188 ( 171 - 190)
    INTEGRAL Likelihood = -3.50    Transmembrane 200 - 216 ( 199 - 221)
    INTEGRAL Likelihood = -2.18    Transmembrane 352 - 368 ( 352 - 371)
40  PERIPHERAL Likelihood = 3.82    286
    modified ALOM score: 2.39

*** Reasoning Step: 3

45  ----- Final Results -----
        bacterial membrane --- Certainty=0.4779 (Affirmative) < success
        bacterial outside --- Certainty=0.0000 (Not Clear) < success
        bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < success

```

50 The protein has homology with the following sequences in the databases:

ORF01622(313 - 1512 of 1812)
 gp|9656920|gb|AA95495.1||AB004305(1 - 418 of 418) NupC family protein [Vibrio cholerae]
 %Match = 24.0
 %Identity = 39.5 %Similarity = 65.7
 Matches = 160 Mismatches = 134 Conservative Sub.s = 106

[illegible]

-2274-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6249> which encodes the amino acid sequence <SEQ ID 6250>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2196 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 211/223 (94%), Positives = 217/223 (96%)

Query: 1 MEVKDILKTVDHLLATTATWFEIQITLDDAMAYETASACIPASVYKKAETVYSGKLAIC 60
+EVKDIKTVDHLLATTATWFEIQITLDDAMAYETASACIPASVYKKAETVYSGKLAIC
Sbjct: 1 VEVKDILKTVDHLLATTATWFEIQITLDDAMAYETASACIPASVYKKAETVYSGKLAIC 60

Query: 61 TVIGFFNGYSTTAAKVFECQDAIQNGADEIDWVNLTDVYKQDFVTVEEIRQKAKCQD 120
TVIGFFNGYSTTAAKVFECQDAIQNGADEIDWVNLTDVYKQDFVTVEEIRQKAKCQD
Sbjct: 61 TVIGFFNGYSTTAAKVFECQDAIQNGADEIDWVNLTDVYKQDFVTVEEIRQKAKCQD 120

Query: 121 HILKVVIVETCOLTKEELIELCGVVTRSGADFIKTSTGFTAGATFEDVEVMKIVGEGVK 180
HILKVVIVETCOLTKEELIELCGVVTRSGADFIKTSTGFTAGATFEDVEVMKIVGEGVK
Sbjct: 121 HILKVVIVETCOLTKEELIELCGVVTRSGADFIKTSTGFTAGATFEDVEVMKIVGEGVK 180

Query: 181 IKAAGGSSLEDARKFIALGASRLGTSRIIKIVKNKVEEGTY 223
IKAAGGSSLEDA+ FIALGASRLGTSRIIKIVKN+ + +Y
Sbjct: 181 IKAAGGSSLEDARKFIALGASRLGTSRIIKIVKNKVEEGTY 223
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2020

A DNA sequence (GBSx2131) was identified in *S.agalactiae* <SEQ ID 6251> which encodes the amino acid sequence <SEQ ID 6252>. This protein is predicted to be phosphopentomutase (deoB). Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0546 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC45496 GB:U80410 phosphopentomutase [Lactococcus lactis subsp.
cremoris]
Identities = 275/408 (67%), Positives = 325/408 (79%), Gaps = 7/408 (1%)

Query: 3 QFDRHLVVLDSVIGGAAPDNDPFVNAQVP-----DGASDTLGHISKTGLVLPVMAKI 56
+P RHLVVLDSVIGGAAPDA+ F N V D SUT+GHIS+ GL VPA+ K+
Sbjct: 4 KFRHLVVLDSVIGGAAPDADKFPNHDVETHEAINDVKSUTIGHISEIRGLDVPNLQL 63

Query: 57 GLGNIPRQALKTVPABRNPSYATKLQEVSLGKDTMTGHWMEINGLNIPTFPDTFVNGFP 116
G GNIPR LKT+PA + P+ Y TKL+E+S GKDTMTGHWMEINGLNI PF T+ G+P
Sbjct: 64 GWNIPRPSPLKTIAPAAQKPAAYVTKLEEISGKDTMTGHWMEINGLNIQTFFPTYPEGYP 123

Query: 117 EDIITKIRDFSGRKVIREANKPYSGTAVIDDFGPRQMETGELIITYTSADPVLQIAAHEDI 176
ED++ KIE+PSGRK+IREANKPYSGTAVI+DFGPRQ+ETGELIITYTSADPVLQIAAHED+
```

-2275-

Sbjct: 124 EDLEKIEEFSGRKIIREANKPYSGTAVIDDGFGRLQETGELIIYTSADPVLQIAAHEDV 183

Query: 177 IFLEELYRICYARSITMERFALL-GRIIARPTVGEFGNFTRTANRHDYAVSPFDVTINL 235
I EELY+ICEY RSIT+E ++ GRIIARPTVGE GNF RT R DYA+SFF +TVL

5 Sbjct: 184 ISREELYKICEYRSITLGSIGIMGRI IARPTVGRAGNFERTIDGRKDYALSPPARTVLE 243

Query: 236 KLDQAGIDITYAVGKINDIFNGSGINHDGHNKNSHGIDTLIKTMGLSEFEKGFSTNLV 295
KL +AGIDITY+VGKI+DIEN G+ +DMGHN ++ G+D L+K M +EF +GFSFTNLV

10 Sbjct: 244 KLYAGIDITYSVGLSDIFATYGVKYDMGHNHDMDGVDRLKAMTKTETPTGFSFTNLV 303

Query: 296 DFDALYGHRRDPHGYRDLCEHFDRLPEIISAMRDKDLLLITADHGNDPTAGDHTREY 355
DFDIA YGHRRD GY + +FD RLPEII AM+ DLL+ITADHGNDP+Y GTDHTREY

15 Sbjct: 304 DFDACYGHRDRDVGYGKAIEDFDGRPLPEIIDAMKEDLLMITADHGNDPSYVGTGHTREY 363

Query: 356 IFLLAYSPSFTQNLIPVGHFADISATVADNFGVDVTAMIGESFLQDLV 403
IFL+ +S SF ++FVGHFADISAT+A+NP V A GESFL LV

15 Sbjct: 364 IFLVIFSKSPKEKVLPGVGHFADISATIAENFVSKVAGTGESFLDALV 411

There is also homology to SEQ ID 2740:

20 Identities = 348/402 (86%), Positives = 374/402 (92%)

Query: 1 MSQFDRIHLVVLDSVGI GAAPDANDFPNAGVFDGASDTLGHISKTVGLAVPNMAKIGLGN 60
MS+P+RIHLVVLDSVGI GAAPDA+ F NAGV D SDTLGHIS+ GL+VPMMAKIGLGN

25 Sbjct: 1 MSKFPRIHLVVLDSVGI GAAPDADKFPNAGVADTSDTLGHISEAGGLSPMAKIGLGN 60

Query: 61 IPRPQALKTVAENPESGYATLQVSLGKDTMIGHWEINGLNITEFPDTPWNGFPEDII 120
I RP LKTVP E+NP+GY TLL+EVSLGKDTMIGHWEINGLNITEFPDTPWNGFPEDII+

25 Sbjct: 61 ISRPILKTVPTEDNPTGYVTLLQVSLGKDTMIGHWEINGLNITEFPDTPWNGFPEDII 120

30 Query: 121 TKIEDPSGRKVIRESANKPYSGTAVIDDGFGPROMETGELIIYTSADPVLQIAAHEDII PLE 180
TKIE+PSGRK+IRESANKPYSGTAVIDDGFGPROMETGELI+YTSADPVLQIAAHEDII+P

30 Sbjct: 121 TKIESPSGRKIIRESANKPYSGTAVIDDGFGPROMETGELIVYTSADPVLQIAAHEDII PVE 180

Query: 181 ELYRICYARSITMERFALLGRIIARPTVGEFGNFTRTANRHDYAVSPFDVTINLQDA 240
ELY+ICEYARSITMERFALLGRIIARPTVGEFGNFTRTANRHDYAVSPFDVTINLQDA

35 Sbjct: 181 ELYKICEYARSITMERFALLGRIIARPTVGEFGNFTRTANRHDYAVSPFDVTINLQDA 240

Query: 241 GIDTYAVGKINDIFNGSGINHDGHNKNSHGIDTLIKTMGLSEFEKGFSTNLVDFDL 300
G+ TYAVGKINDIFNGSGI +DMGHNKNSHGIDTLIKT+ L EF KGFSFTNLVDFDL

40 Sbjct: 241 GVPTYAVGKINDIFNGSGITNDMGHNKNSHGIDTLIKTLQLEPFTKGFSFTNLVDFDAN 300

Query: 301 YGHRDDPHGYRDLCEHFDRLPEIISAMRDKDLLLITADHGNDPTAGDHTREYIPLLA 360
+GHRDDP GYRDLCEHFD RLPEII+ M++ DLLLITADHGNDPTAGDHTREYIPLLA

40 Sbjct: 301 FGHRDDPHGYRDLCEHFNRLPEIIANMKEDLLLITADHGNDPTAGDHTREYIPLLA 360

45 Query: 361 YSPSFTQNLIPVGHFADISATVADNFGVDVTAMIGESFLQDL 402
YS SFTQNLIP GHFADISATVA+NGVDVTAMIGESFL L

45 Sbjct: 361 YVSPSFTQNLIPQGHFADISATVAENFGVDVTAMIGESFLSHL 402

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2021

A DNA sequence (GBSx2132) was identified in *S.agalactiae* <SEQ ID 6253> which encodes the amino acid sequence <SEQ ID 6254>. Analysis of this protein sequence reveals the following:

55 Possible site: 35
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -12.05 Transmembrane 9 - 25 (4 - 35)

60 ----- Final Results -----
bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-2276-

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6255> which encodes the amino acid sequence <SEQ ID 6256>. Analysis of this protein sequence reveals the following:

5 Possible site: 56
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -5.57 Transmembrane 41 - 57 (38 - 60)
 10 ----- Final Results -----
 bacterial membrane --- Certainty=0.3230(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9143> which encodes the amino acid sequence <SEQ ID 9144>. Analysis of this protein sequence reveals the following:

15 Possible cleavage site: 49
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -5.57 Transmembrane 13 - 29 (10 - 32)
 20 ----- Final Results -----
 bacterial membrane --- Certainty= 0.323(Affirmative) < succ>
 bacterial outside --- Certainty= 0.000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

25 An alignment of the GAS and GBS proteins is shown below.

Identities = 276/544 (50%), Positives = 368/544 (66%), Gaps = 5/544 (0%)

Query: 5 FKKKVKVCLVIFGIVLVSLLSGFFYFSKGQVLSRFVAARSRTSGQAFINIKRYVMVSD 64
 F K +K +I L L G FY+SK ++ ++ ARS SG F+NIK Y+VM D
 30 Sbjct: 33 FHHKKLQITIIAATSLFLFLIGAFYYSKHNHCINAYLKARSAQSSPFVENIKAYLVNDD 92

Query: 65 TGESITNDEANYANFPEPLSKSEARKLQGEIKENQNDMSYLFKRVGSLRGLFPDYRIANKP 124
 T E ITNDEA Y F S+ E R+ Q++K +++ ++ +K VG R IFPDYRIA KP
 35 Sbjct: 93 TNEQITNDEMYTKFRYSQKELRQKQDLGAASQDSAVQVSVGRRFWLFPDYRIAIKP 152

Query: 125 MSLLTKTNVPKLDVLLNQKCVATSNDSHFVSTVERLEPRTHTYASLEGTSDGKSEIKKDY 184
 M LT+KTNVP+ DVLNQKQVA S+S+ FSF +R+LP YTAS+ G +G+ IK+ K Y
 40 Sbjct: 153 MDLTKTNVPQADVLLNQKQVAVSDSBSQSVKLDRLPTAETASIRGRHGRNKNKNSY 212

Query: 185 DGNKNTIDLSVAFKSPFTVTSNLMQNLPGDNRIAKLDKDGSHVENYFVTDGSKAYIKQV 244
 DG N -DLSV+P+P VTSN G+LYF IN I LKDG VE+YFVT+ ++AY+K
 45 Sbjct: 213 DGNPVLDSVSFRTFLVTSNAKQGDLYPDNHIQTLKDGQLQVEDYPTVENPAQAYMKT 272

Query: 245 FNDGEITSHKQKLISADNQTIKLDVGLINEKPAQKLIKITAPOQLILYVSTGQDPQTLG 304
 F.DGE+ S K L + + T+++ V LL E +AG+ L+AF+QL+ Y+STGD L
 50 Sbjct: 273 FPDQRLRSQKALADVEGATLEILLVTDLLSEDAKAGELLVSAPDQIMHLYSTGQDSNR 332

Query: 305 TVFKEAGENDFYKGLKESIKAKFPVDNDRKASHFTINIVLNMTQVQKESYQVNFADYD 364
 +VFE G+ N FY+GLKESIKAKF TD RKAS IP+I+L MTQVKG +Y +F+ A Y+
 55 Sbjct: 333 SVFEAGSSNAPFYRLKESIKAKFPQDTRKASRLINIPILLTMTQVKGHTTVLDPTATYE 392

Query: 365 FNYDKSTDPKTKTYGHIQNLITGNFIMKSSGNSYLISNDGKKDITVAKETINKVADPVI 424
 F YDKSTDP+ T GHI Q+LITG +KK G YLIS G K+ITV KE N++K S+
 60 Sbjct: 393 FLYDKSTDPQBSTGHIHQDLTKVTVKVKVQGHYLSQSGSKNITVYKDNQLQAP--SV 450

Query: 425 FPNLAVGSGWKEGVEDGTVTMTFVDKQKVTKR-KVYKSKSKSNHSAKVTKLEKQNGLY 483
 FPE++G+M G+ ++ M+ DG +T K + K ++SKE+ +K++K+EDKNG Y
 65 Sbjct: 451 FPESILGTWTGQANGLSHMSLASDGTITTKVEDQKGNRSKET-RTAKISKVEDKNGNPF 509

Query: 484 LYQYESGTDITTFV-TGGIGGLKVYKYGIKIEGNKIIPVIMQTSDDGDFDYHKLPLSKP 542
 LY + G+D + V GG+GG VKYANG KI G PV+WQ + EFDY KPL
 70 Sbjct: 510 LYTDPGSDISALVPBGGLGAGNVKYGFKLSGKTASPVVWQALTHEFDYTKPLSGVT 569

-2277-

Query: 543 LTQ 546
 L QQ
 Sbjct: 570 LQK 573

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 9065> which encodes amino acid sequence <SEQ ID 9066>. An alignment of the GAS and GBS sequences follows:

Score = 47.3 bits (110), Expect = 4e-07
 Identities = 65/303 (21%), Positives = 119/303 (38%), Gaps = 18/303 (5%)

Query: 153 FYILGIGTSISIVALTAFVKHISLNFKSIKKLANRKGIEVLSENHNSYQII---SFDDI 209
 +YIL + T I+ +V + +S F +KKL KM + +QI EF D+
 Sbjct: 37 YYILSV-TILACTVGGIVNLFLLSSVFTSLKKLQKMKDISQRCPTDKAQICSQDFNKL 95

Query: 210 LRTLHAKGDMKLSLIEREILEKQDLSPQIAALSHDIKTXXXXXXXKXXXXXXXXXQ 269
 + L+S + +++ + IA LSHDIKT
 Sbjct: 96 ETAFNQMSSELESTFKSLNBSREKTMIAQLSHDIKTPTSISQSTVEGILDGIISEREV 155

Query: 270 GYIVSMNNSISVFEGYFNSLISYTRML-----SEDRSVKILVBEILLSLHFEVDL 321
 Y + N+IS N L+ + +E + I +++LL ++ E +
 Sbjct: 156 NYTL---NTISRQTNLNLHLVEELSFTLETMSDTAEFHKETTYLDKLLIDILSEFQLV 212

Query: 322 LWINNIEPSICNRLIITSFYGDERNLIRALSNLLVNAIRFMPVLDDKKIEVI--SESGQIH 381
 N + I ++ + L R L NL+ NA ++ + + + I
 Sbjct: 213 FEKENRQVHVDVAPDVSKLSSQDKLSRILLNLISNAKKYSDP-GSPLTIKAYNRQDIV 271

Query: 382 FEIWRNGERFSDSTLKKDKLFTYEDYSRGNK--HYGIGLAFVKGVALKHGNLQLNPA 439
 +I + G D L Y + SR K +G+GL + +A + G++ + +
 Sbjct: 272 IDIDQGYIKDEDLASIFNRLYRVSSRNMYGGLGLYTAQALHQLNDILVESQY 331

Query: 440 RGG 442
 + G
 Sbjct: 332 QKG 334

A related sequence was also identified in GAS <SEQ ID 9135> which encodes the amino acid sequence <SEQ ID 9136>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -3.56 Transmembrane 145 - 161 (145 - 164)

----- Final Results -----

bacterial membrane --- Certainty=0.2423 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

SEQ ID 6254 (GBS280) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 8; MW 63.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 7; MW 88.7kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2022

A DNA sequence (GBSx2133) was identified in *S.agalactiae* <SEQ ID 6257> which encodes the amino acid sequence <SEQ ID 6258>. This protein is predicted to be ribosomal large subunit pseudouridine synthase D (rluC). Analysis of this protein sequence reveals the following:

Possible site: 22

-2278-

>>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -4.62 Transmembrane 2 - 18 (1 - 19)

----- Final Results -----

- 5 bacterial membrane --- Certainty=0.2848(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 10 >GP:CAB12749 GB:Z99108 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 97/251 (38%), Positives = 147/251 (57%), Gaps = 15/251 (5%)

Query: 86 KHVLLNNEFINMQTVVQENDTITLIFDDEYPTKKIPLGRARLIDCLYEDEHLIIVNKPE 145
 ++ N+E + ++ D + + + + G +D L+ED H++I+HKP

15 Sbjct: 43 QQ:KVNHSVLNNMIVKGEKRVFIDLQSEASSVPEYGE---LDLIFEDNHLIINKPA 99

Query: 146 GMKTHNQPNBIALNHSVAY---SQQTCTV--VHRLDMETSGAVLFAKNPFILFLINQ 199

G+ TH N+ + L ++ AY +G+TC V VHRLD +TSGA++FAK+ +++Q

20 Sbjct: 100 GIATHPNEDGQTGLANLIAYFYQINGETCKVRHVHRLDQDTSGAIVFAKHLRAHAILDQ 159

Query: 200 RLERKEIMREYVALVEGKFSPEKQVLARDKIGRNR-HDRRKRIIDSKNGQHMTIIDVL-- 256

+LE+K + R Y A+ BGE K + IGR+R H R+R+ S QG A+T V+

Sbjct: 160 QLEKKTLLKRTYTAIEAGKELRTKGTINPPIGDRDSHPTRRV--SPGGQTAVTHFKVAVAS 217

25 Query: 257 KYIQNSSLIKRCLETGRTHQIRVHLSHHKHPGLIGDPLYPSSN-NERIMLHAHRLTLSHP 315

+ SL++ LETGRTHQIRVHL+ GHPL GD LY S R LHA++H NP

Sbjct: 218 NAKERLSLVELBLSTGRTHQIRVHLASLGHPLTGDSLYGGSSKLNRQALHANKVQAVHP 277

30 Query: 316 LTCETISVEAP 326

+T E I EAP

Sbjct: 278 ITDELIVAEAP 288

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6259> which encodes the amino acid sequence <SEQ ID 6260>. Analysis of this protein sequence reveals the following:

- 35 Possible site: 38
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 40 bacterial cytoplasm --- Certainty=0.4198(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 172/278 (61%), Positives = 212/278 (75%), Gaps = 2/278 (0%)

45 Query: 63 TVKELLEDFLIPKIRHFLRVKKHVLNNEFINMQTVVQENDTITLIFDDEYPTKKIP 122

TVK LLE+ LIPKIRHFLR KKHVLN +NQ+ V+ D + L PD EDYP K I

Sbjct: 2 TVKALLEBQLLIPKIRHFLRTKKHVLNHSVNNQSCVKYGDQVKLFPDHDYPEKIV 61

50 Query: 123 IGRARLIDCLYEDEHLIIVNKPBMKTHNQPNBIALNHSVAYSQQTCTVVHRLDMETS 182

+G+AE + CLYEDEH+IIVNKPBMKTHN P E+ALLNHSVAY+QQTCTVVHRLD ETS

Sbjct: 62 MQQAKVTCLYEDRHIIIVNKPBMKTHNDPTRIALLNHSVAYTQCTVYVHRLKETS 121

55 Query: 183 GAVLFAKNPFILFLINQRLERKEIMREYVALVEGKFSPEKQVLARDKIGRNRHRRKRIID 242

GA+LEAK PFIL++N+ LE++I RSY ALV G IGR+RHRRKRR++D

Sbjct: 122 GAILFAKTYPFILFLINRLERKRTHREYALVHGSLSDFVTVTHEPIGRHRRHRRKRVVD 181

60 Query: 243 SKNGQHMTIIDVLK-YIQNSSLIKRCLETGRTHQIRVHLSHHKHPGLIGDPLY-PPSSN 300

NQ+ A+T + ++K + + +SL+ C+L+TGRTHQIRVHL+H GH L GDPLY N +

Sbjct: 182 PINGKKAITEVTLVKNFHTASLTQLQCTGRTHQIRVHLAQGHVLPDGLYSNGKDC 241

Query: 301 ERIMLHAHRLT+SHDILTCTETISVEAPSTFFKILNNYK 338

RIMLHA++L L HPLT E I V+A S+TF+ +LN K

Sbjct: 242 ARIMLHAQLRLKHPLTQEDICVQAKSATFDVAINAQK 279

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2023

- 5 A DNA sequence (GBSx2134) was identified in *S. agalactiae* <SEQ ID 6261> which encodes the amino acid sequence <SEQ ID 6262>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.02 Transmembrane 98 - 114 (93 - 119)
 10 ----- Final Results -----
 bacterial membrane --- Certainty=0.4609 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 15 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF04735 GB:AF101780 penicillin-binding protein 2a
 [Streptococcus pneumoniae]
 Identities = 424/773 (54%), Positives = 555/773 (70%), Gaps = 47/773 (6%)
 20 Query: 2 KLDFKFDLFRVDENDSEMRKNEQSTRSETSNLDGEEVYDIDITRPSKSOYQYRGIRHQ 61
 KLDF+KF+ LF+ +ETS L+ + I R S+S
 Sbjct: 5 KLFEKFLSLFK-----KETSLESDSD---STLLRRSR----- 34
 25 Query: 62 KENAKSRPEMLQKVDRYLSPKPNPIRRPWRRYRIGKLLFIAMAFILIFGSYLFYLSKTA 121
 DR + PIR+PWRRY+ K++ I ++ L+ G YLF ++K+
 Sbjct: 35 -----DRKKLAQVGPIRPFWRKYHLTKTILILGLSAGLVGIYLFPAVAKST 80
 30 Query: 122 TVSDLSQALKTTTTTYDRNKEYAGLSQCKGTIVELNAISHLNNAVIATEDRTPYENNG 181
 V+DLQ+ALKT T I+D+ ++ AG LSGCKGTIVEL IS +L+NAVIATEDR+FY+N+G
 Sbjct: 81 NVNDLQNALKTRTLIFOREEKEAGALSQCKGTIVELTDISENLQNAVIATEDRSFYKNDG 140
 35 Query: 182 VNFKRFFLAVALTKFGGGSTITQQLANAYLSQDQTKRKAREFFLALELTKYKSAEI 241
 +N+ RFFLA+ T G+ GGGSTITQQLANAYLSQDQ++RKA+EFFLALEL+KYSK +I
 Sbjct: 141 INYGRFFLAIVTAGRSGGGSTITQQLANAYLSQDQTVRKAREFFLALELSKYSKSKQI 200
 40 Query: 242 LTMVLNYSFGNGVGVGEDASRKYPOTSANLTVDEAATLAGMLKGPEVYNPPYSVENAT 301
 LTMVLN+YFGNGVGVGEDAS+KYFG SA+ +++D+AATLAGMLKGPE+YNP SVE++T
 Sbjct: 201 LTMVLNAYFGNGVGVGEDASKYFGVBAEVSVDQAATLAGMLKGPYLNPLNSVEDST 260
 45 Query: 302 NRRDTVLQNVMDAGKLTKSQAKESAGIMQRLADTVAGKINDYRYPSPYDAVNEAIDT 361
 NRRDTVL NV AG + K+Q EAA + M ++L D Y GKI+DYRYPSPYDAVNEA+
 Sbjct: 261 NRRDTVLQNVVAAGYIDKNQTEAALVEMTSQLEHDKYBKISDYRYPSPYDAVNEAVSK 320
 50 Query: 362 YGISEKDIVNNGYKIYITLQDQNYQSGMKTFFDTSLPFVSDYDQQAQASVALNDPKTGG 421
 Y ++E++IWNNGY+IYT LDQNYQ+ MQ ++TSLFP ++ DG AQ SVAL+PKTGG
 Sbjct: 321 YNLTEERIVNNGYRIYITLQDQNYQNMQIYVENTSLFPRAE-DGTVAQSGSVLEPKTGG 379
 55 Query: 422 VRGLVGRVSTKDAQFRSPNYATQSKRSPASTIKPLVVYSGPAISGWSITKELNIVQDF 481
 VRG+VG+V FR+FNATQSKRSP STIKPLVVY+FA+ G++K+L N +
 Sbjct: 380 VRGVVGVQVANDICGFRNPNYATQSKRSPGSTIKPLVVYTPAVAGKALNKQLNHTMQY 439
 Query: 482 HGVKPSNYGGIET-ESLIMVQALANSYNIPAVYTLKLGINKAFTYGRKPOLANSSANKE 540
 YK NY GI+T +PMVQ+LA S N+EV T++ LG++KAF G KGLAM ++
 60 Sbjct: 440 DSYKVDNYAGIKTSREVPYQSLASLMLPAVATVNDLGVNKAPEAGEKPGIAMEKVDVR 499
 Query: 541 LGVALGSGVITNPLQMAQYAS+TANDGIMRAHLITRIETANDKLVQPTDKPKRVISRS 600
 LGVALG V TNPL+MAQAY+ PAN+G+M AH I+RIE A+G+++ + KRVI +
 Sbjct: 500 LGVALGSGVETNPLQMAQYAA+PANGELMRAHPIRIRIENASQVIAHSKNSQKRVIDS 559
 65 Query: 601 VASKNTSNMLGTFSNGTALDANVYGYTMAGKIGTITDPSNLSQDQVGVYGTEDVVTSQ 660
 VA KNTSNMLGTF+NOT I+++ Y MAGKIGTITE PNP + DQNV+GYTDEDVVIS

-2280-

Subjct: 560 VADKMTSMMLGTFITNGTGISSSPADYVMAGKTGTTTAVFNPEYTSQDWVIGYTPDGVISH 619

Query: 661 WVGFIQYTDKHHYITDSSAGTASNIFSTQASYILPYTKSSSFTHIENAYFQNGISGVYNAQ 720

W+GF TD++HYL S++ A+++F A+ ILPTT GS+FT+ENAY QNGI +

Subjct: 620 WLGFTTDEHYLAGSTNGAAHVFNRIANTILPYTPGSTFT-VEAYKQNGIAPANTKR 678

Query: 721 DASNTTQESRSIINDLKDSASKAAQDISRAVEDSNPQKVKWDANWSLKYDF 773

N ++ ++D++ A + SRA+ D+ +EK + W+S+ +FR

Subjct: 679 QVQTDNDSQTDNLSDIRGRAGSLVDASRAISDAKI KEKAQTWDSIVNLF 731

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6263> which encodes the amino acid sequence <SEQ ID 6264>. Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -7.96 Transmembrane 104 - 120 (99 - 124)

----- Final Results -----

bacterial membrane --- Certainty=0.4185(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AA04735 GB:AF101780 penicillin-binding protein 2a [*Streptococcus pneumoniae*]
 Identities = 414/730 (56%), Positives = 539/730 (73%), Gaps = 17/730 (2%)

Query: 50 TKNSBQDPATLQSRAYEGSPKSRPAWLQLEAVLPSPQRPPIRRFWRHYHGLKMLILI 109

T E +T L+RSR+ +XL V PIR+FWRYH+ K++IL

Subjct: 18 TSELESDSTILRRSRSDR-----KKLAQV-----GPIRKFWRYHILTKIILIG 62

Query: 110 GTLVLLGSYLFYLSKTAKVSDLDQALKATTIVYDHKGEYAGLSQKSGSYVELNATSD 169

+ LL+G YLF +K+ V+DLQ+ALK T+I+D + + AG+LSGQKG+YVEL IS +

Subjct: 63 LSGLLVGIYLFVAVKSTVNDLQNALKTRTLIFOREEKBAGALSGQKGTIVELDTISN 122

Query: 170 LENAVTATEDRTFYNSGDNLRKFLAVVTAAGFGQGGSTITTQQLANAYLSQDQTKRKA 239

L+NAVITATEDR+FY N GIN RF LA+VTAGR GGGSTITTQQLANAYLSQDQ++RKA

Subjct: 123 LQNAVITATEDRSFYKNDGINYGRFFLAIVTAGRGGGGSTITTQQLANAYLSQDQTVRKA 182

Query: 230 REFFLALELTKYKSKDILTYMLNNSYFGNGVNGVEDASQKYFGTTAANLTDBAATLAG 289

+EFFLALEL+KKYK+ ILTYMLN+YFGNGVNGVEDAS+KYPG +A+ ++LD+AAITLAG

Subjct: 183 KEFFLALELKKYKSKDILTYMLNAYFGNGVNGVEDASKYFVGSSEVLDQAATLAG 242

Query: 290 MLKGPEIYNPYHSLKNAATHRRDVTLGAMVDAKKI TQYAAQARAVGLKRLADTYVGKD 349

MLKGPE+YNP +S++++T+RRDVTLV MV A I + + +A V + ++L D Y GK

Subjct: 243 MLKGPEIYNPIANEVEDSTNRRDVTLVQNMVAAGTIDNQETEAAEVDMTSGLEHDKYEGKIS 302

Query: 350 DYKYPSYFDAVISRAIATYGLSEKDIVNNGYKYVYELTDQNYQ+QWQYTPANDELFPVSAY 409

DY+YPSYFDVA++EA++ Y L+E++IVNNGY++YELTDQNYQ MQ + N LFP A

Subjct: 303 DYKYPSYFDVAVNEARVKNYLTBBEIVNNGYRIYELTDQNYQANQIYIYENTSLFP+RAE 361

Query: 410 DGSSAAQASVALDPKTKGVKRLIGKVNSSSEPFTRSPNYATQAKRSASTIKLVVYPA 469

DG+ AQ+ SVAL+PKTKGVGRG++G+V ++ PK+FNATY+KRSF STIKLVVY PA

Subjct: 362 DGTFAQSGSVALEPKTKGVGRGVGQVADNDKTPGNPNYATQSKRSPSTIKLVVYTPA 421

Query: 470 VASGWSIEKELPNTVQPDGQPHNY -GNYESDVMYQALANSYNI PAVSTLNDIGDK 528

V+GW++ K+L N +D Y+ NY G S +VVMYQ+LA S +PAV+T+ND+G+DK

Subjct: 422 VEAGWALNKQLNHTMQTDSYKVDNYAGI KTSREVMYQSAEASLNPATVNDLGVDX 481

Query: 529 AFTYGTGTPGLEMSAKKELGVALGSGSVTTNPLEMAQAYAAFNNGVIHFAHLNRIENAR 588

AF G+ PGL+M + LGVALS V TNPL+MAQAYAAAFAN G++ AG I+RIENR

Subjct: 482 AFEAGKPGKLANMEKRVRLGVALGSGSVGTNPLQMAQAYAAAFANGLMPEAHFI SRILENS 541

Query: 589 GEVLKFTFDKAKRVVSGSADKMTAMMLGTFPSNQTAVNANVYGTTLAGKTGTITTFNPN 648

G+V+ + + KR+ +SVADMT+MMLGTF+NST+ +++ Y +AGKGTGTE PNP+

Subjct: 542 GQVITASHKNSQKRVIDKSVADKMTSMMLGTFITNGTGISSSPADYVMAGKTGTTAVFNPE 601

-2281-

Query: 649 LAGDQWVIGYTPDQVVISQWVGPNQDGNHYLTDSSAGTASAIPTQASYILPYTKGSQFH 708
 LQWVIGYTPDQVVIS W+GF TDENHYL S++ A+ +F A+ ILPTT GS F
 Sb|ct: 602 YTSLQWVIGYTPDQVVISHWLGFPTTDENHYLAGSTSNAAHVFRNIAANTILPYTPGSTFT 661

Query: 709 VDNAYACNGISAVYGVNETGNQSGVDITQSIIDGLRKSQAQASQSLSKAVDQSGLRDKAQ 768
 V+MAY QNGI+ + T + +R AQ S+R+ + ++KQ+
 Sb|ct: 662 VDAYKNGIAIPANKRQVQVINDNSQTDNLSDIRGAQSLVDASRAISDAKIKKKAQT 721

Query: 769 IWKEIVDYFR 778
 IW IV+ FR
 Sb|ct: 722 IWSIVNLFR 731

An alignment of the GAS and GBS proteins is shown below.

Identities = 530/715 (74%), Positives = 623/715 (87%), Gaps = 1/715 (0%)

Query: 59 RHQKENAKSRPFWLQKVDVRLPSPKPIRRPWRRYRIGKLLFIALMAFILIYGSYLFYLS 118
 R + + KSRP WQK++ LPSP+ PIRRPWRRY IGKIL I + +L+ GSYLFYLS
 Sb|ct: 65 RAYEGSPKSRPAWLQKLEAVLSPQRPPIRRPWRRYRIGKLLMILIGTIVLLGYSYLFYLS 124

Query: 119 KTATVSDLQSAKLTITTIYDKNKFYAGKLSGQKGTIVYVSLNAISDHLQNAVIATEDRTFYE 178
 KTA VSDLQ ALK TT IYD SYAG LSGQKG+YVELNAISD L+NAVIATEDRTFY
 Sb|ct: 125 KTAKVSDLQDALIKATTVIYDHKGVSAGLSGQKGTIVYVSLNAISDLENAVIATEDRTFYS 184

Query: 179 NNGVNFRFFLAVATLKGFGGGSTITQQLAKNAYLSQDQTIKKAREFFLALELTKKYSK 238
 N+G+N KRF LAV T G+FGGGSTITQQLAKNAYLSQDQTIKKAREFFLALELTKKYSK
 Sb|ct: 185 NSGINLKRFLAVATVGRFGGGSTITQQLAKNAYLSQDQTIKKAREFFLALELTKKYSK 244

Query: 239 AEILTWYLNNSYFGNGVGVVEDASRKYPGTSAANLTVDENATLQMLKGPEVYNFYYSVE 298
 +ILTWYLNNSYFGNGVGVVEDAS+KYFGT+AAANT+DEAATLQMLKGPE+YNYF+S++
 Sb|ct: 245 KDILTWYLNNSYFGNGVGVVEDASQKYPGT+AAANTLDEAATLQMLKGPEYINPYHSLK 304

Query: 299 NATNRDITVLAAMVDAGKLTKSQAKEASIGMHRRLADTYGKINDYRPSYFDVAVNEA 358
 NAT+RRDITVL AMVDA K+T++A++A++G+KRLADTY GK +DY+PSYFDVAV+EA
 Sb|ct: 305 NATHRRITVLAGAMVDAKKTITQ+KACQARAVGLKRLADTYGKTDYKYPSYFDVAVSEA 364

Query: 359 IDTYGISEKDIVNNGYKIYITALDQNYQSGMQKTFDDTSLFFVSDYDQSAQASVALDPK 418
 I TYG+SEKDIVNNGYK+YT LQNYQ+GMQ TP++ LFFVS YDG SAQ ASVALDPK
 Sb|ct: 365 IATYGLSEKDIVNNGYKYVTELDQNYQTMQQTTFNDELFPVSAZDSSAQAASVALDPK 424

Query: 419 TGGVRLGVRGVQSTKDAQFRSPNYATQSKRSPASTIKPLVYVSPAIASGNSIDKELPNKY 478
 TGGVRLG+GRV S+++ FRSPNYATQ+KRSPASTIKPLVYV+PA+ASGNSI+KELPN V
 Sb|ct: 425 TGGVRLGIRGVNSSENPTFRSPNYATQAKRSPASTIKPLVYVAPAVASGNSIBEKELPTV 484

Query: 479 QDFHGYKPSNYGSIETESIPMYQALANSYNIPAVYTLQKLGINKAFTYGRKGLNMSAN 538
 QDF GY+P NYG E+B +PMYQALANSYNIPAV TL++ +GI+KAPTYG+ EGL+MSSA
 Sb|ct: 485 QDFDGYQPHNYGNYSESDVPMYQALANSYNIPAVTINDIGIDKAPTYGKTFGLDMSAK 544

Query: 539 KELGVALGGSVTTINPLEMAQAYSTFANDGINHRAHLITRISTANGKLVQVQTDKPKRVIS 598
 KELGVALGGSVTTINPLEMAQAY+ PAN+G++H AHLI RIE A G+++K PIDK KRV+V
 Sb|ct: 545 KELGVALGGSVTTINPLEMAQAYAAFAFANNGVTHPAHLINRIENARGEVLKTFDIAKRVVS 604

Query: 599 RSVASEMTSMMLQTFNSGTAINANVYGYTMAGKTGTTETDFFNPLNSGDCWVGYYTPDQVVI 658
 +SVA RMT+MMMLQTFNSGTA+ANANVYGYT+AGKTGTET+RNP+L+GDQWV+GYTPDQVVI
 Sb|ct: 605 QSVADMTAMMLQTFNSGTAINANVYGYTLAGKTGTETTFNENELAGDQWVGYTPDQVVI 664

Query: 659 SQWVGFKMTDKHNYLTDSSAGTASAINFTQASYILPYTKGSSFTHIENAYFCNGIGSVYN 718
 SQWVG F TD++HYLTDSSAGTAS IFTQASYILPYTKGS F H++NAY QNGI +YY
 Sb|ct: 665 SQWVGFKMTDKHNYLTDSSAGTASAIPTQASYILPYTKGSQF -HVNAYACNGISAVYG 723

Query: 719 AQDASNTTNCSSRSIINDLAKDSAKAAQDISRAVEDSNPQEKVKDAMNSLKDYFR 773
 + N+ + ++SII+ L+ SA +A+Q +S+AV+ S ++K+ W + DYFR
 Sb|ct: 724 VNETGMQSGVDITQSIIDGLRKSQAQASQSLSKAVDQSGLRDKAQSIWKIVDYFR 778

SEQ ID 6262 (GBS397d) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 153 (lane 13; MW 76kDa) and in Figure 184 (lane 9; MW 76kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2024

A DNA sequence (GBSx2135) was identified in *S. agalactiae* <SEQ ID 6265> which encodes the amino acid sequence <SEQ ID 6266>. This protein is predicted to be M-like protein. Analysis of this protein sequence reveals the following:

```

Possible site: 27
>>> Seems to have no N-terminal signal sequence
10  INTEGRAL Likelihood = -10.56 Transmembrane 609 - 625 ( 599 - 628)
    INTEGRAL Likelihood = -0.00 Transmembrane 19 - 35 ( 19 - 35)

---- Final Results ----
15  bacterial membrane --- Certainty=0.5225(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GF:CB91647 GB:AJL30830 cell wall protein, putative [Zea mays]
Identities = 106/182 (58%), Positives = 123/182 (67%), Gaps = 8/182 (4%)

20
Query: 396 KEDKDKVQPEAKPEAK--PDVQPEAKPDVQPEAKPDVQPEAKPDVQPEAKPDV--KPEA 451
      K + KP+ KPE KPE K P KPE KP+ KPE KP+ KPE KP KPE KP+ KPE
Sbjct: 116 KPEPKPEPKPEPKPEPKFKPKPKPEKPEPKPEKPEPKPEKPKPKPEKPEKPEKPE 175

25
Query: 452 KPDVQPKPAKPDVQPEAK--KPDVKPDVQPKVQPEA--KPEDKPDVQPKVQPEAKPDVQKFE 507
      KP+ KP+ KP+ KPE KP+ KP+ KP+ KPE KP+ KP+ KP+ KPE KP+ KPE
Sbjct: 176 KPEPKPEPKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPE 235

30
Query: 508 KPEAKPEKPEAKPEAKPEAKPDVKPEAKPDVKPEAKPEAKPEAKPEAKPEAKPEAKPEAK 567
      KPE KPE +PE KPE KPE KP D +P KPE KDE KDE K E KPE K E KPE
Sbjct: 236 KPEPKPEPKPEKPEKPEKPEKPKPKPKPKPKPKPKPEKPEKPEKPEKPEKPEKPE 295

Query: 568 KP 569
      KP
35
Sbjct: 296 KP 297

```

There is also homology to SEO ID 822.

A related GBS gene <SEQ ID 8957> and protein <SEQ ID 8958> were also identified. Analysis of this protein sequence reveals the following:

```

40 Lipop: Possible site: -1  Crend: 8
   MoG: Discrim Score: -5.20
   GVH: Signal Score (-7.5): 3.07
       Possible site: 27
   >>> Seems to have no N-terminal signal sequence
45 ALOM program count: 2 value: -10.56 threshold: 0.0
   INTEGRAL Likelihood =-10.56 Transmembrane 609 - 625 ( 599 - 629)
   INTEGRAL Likelihood = -0.00 Transmembrane 19 - 35 ( 19 - 35)
   PERIPHERAL Likelihood = 8.54 139
   modified ALOM score: 2.61

50 *** Reasoning Step: 3

----- Final Results -----
      bacterial membrane --- Certainty=0.5225 (Affirmative) < succ-
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ-
55      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ-

```

LPX'TG motif: 596-600

-2284-

Example 2025

A DNA sequence (GBSx2136) was identified in *S.agalactiae* <SEQ ID 6267> which encodes the amino acid sequence <SEQ ID 6268>. This protein is predicted to be transcription antitermination protein nusG (nusG). Analysis of this protein sequence reveals the following:

```

5   Possible site: 48
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10  bacterial cytoplasm --- Certainty=0.3203 (Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP: CRA53738 GB:X76134 nusG [Staphylococcus carnosus]
    Identities = 90/175 (51%), Positives = 118/175 (67%), Gaps = 2/175 (1%)

    Query: 7   KGWFLVLTQYSGYENKVKENILQRAQTNNMLDNILRVETPTQTVNVEKNKGKTEIENRFP 66
              K W+ + T YSGYENKVK+NL +R ++ NM + I RV IP + K+GK K++ + PP
    Sbjct: 8   KRWTAVHTYSGYENKVKKNLEKRVESNNNTBQIPRVVLPBEEETQVKDGGKAKLTKTKPT 67

20  Query: 67   GTVLVEVMVWIDEAMFVVRNTFNVTGTFVSGHGNRSKPTPLLEEIRSLISMGQTVDFDPT 126
              GTVLVE+VNTDB+W+VVRNTP VTGTFVGS G SKP PLL +E+R IL M3 D
    Sbjct: 68   GTVLVELVNTDBSWVVRNTFGVTGTFVSGAGAGSKPNPLLPDEVRFLKMGKMKETIDV 127

25  Query: 127  NIKEGDUVQIIDGAFIQGBGRVVEIENKVKVL--MINMFGSETQAEILEYQVAEL 179
              ++ G+ V+I G F Q G V EIE +K KL +++MFG ET E+E Q+ +L
    Sbjct: 128  EVEVGQVRIKSGFFANQVGEVQIEADKFLTVLVDNMFGRETPVEVEFDQIEKL 182

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6269> which encodes the amino acid sequence <SEQ ID 6270>. Analysis of this protein sequence reveals the following:

```

    Possible site: 53
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
35  bacterial cytoplasm --- Certainty=0.3874 (Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

40  Identities = 170/179 (94%), Positives = 178/179 (98%)

    Query: 1   MLDSPDKGFWFLVLTQYSGYENKVKENILQRAQTNNMLDNILRVETPTQTVNVEKNKGKTEI 60
              MLDSPDKGFWFLVLTQYSGYENKVKENILQRAQTNNMLDNILRVETPTQTVNVEKNKG+TKET
    Sbjct: 6   MLDSPDKGFWFLVLTQYSGYENKVKENILQRAQTNNMLDNILRVETPTQTVNVEKNKGQTKET 65

45  Query: 61   EENRFPGYVLVSWVWIDEAMFVVRNTFNVTGTFVSGHGNRSKPTPLLEEIRSLISMGQT 120
              EENRFPGYVLVSWVWIDEAMFVVRNTFNVTGTFVSGHGNRSKPTPLLEEIR+IL+SMGQT
    Sbjct: 66   EENRFPGYVLVSWVWIDEAMFVVRNTFNVTGTFVSGHGNRSKPTPLLEEIRALLISMGQT 125

50  Query: 121  VDVFDTNIKEGDUVQIIDGAFIQGBGRVVEIENKVKVLMINMFGSETQAEILEYQVAEL 179
              +DVFDTNIKEGDUVQIIDGAF+GQGBGRVVEIENKVKVLM+NMFGSET AE+ELYQ+AEI
    Sbjct: 126  IDVFDTNIKEGDUVQIIDGAFIQGBGRVVEIENKVKVLMINMFGSETVAEVELYQIAEL 184

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2285-

Example 2026

A DNA sequence (GBSx2137) was identified in *Sagalactiae* <SEQ ID 6271> which encodes the amino acid sequence <SEQ ID 6272>. This protein is predicted to be a glycosyl transferase. Analysis of this protein sequence reveals the following:

5 Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 10 bacterial cytoplasm --- Certainty=0.1558 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAF228363 GB:AF224467 putative glycosyl transferase [Haemophilus
 ducreyi]
 Identities = 98/259 (37%), Positives = 155/259 (59%), Gaps = 10/259 (3%)

Query: 5 VALAVDSNYLDKALVTIKSICVYNNRNTFYLPKQDTPEVWVRNIRKLEPLGSKLINVKI 64
 + LA + +Y + L TIKSI ++N++I FYL N+D P EW +N KL L S++I++K+
 20 Sbjct: 10 IVLAAQDSYSEYILTITKISYILRNKHIRFYLLNRDYPTEWFDILNNKLKRLNSELIDIKV 69

Query: 65 VNYDIAHLTFTFLTVS---TWFLFLADYIPSSRVLYLDSDIIVTNNLDYLFELDFKGYLL 121
 N I + T+ +S T+FR F++D+I +V+YLD+DI+VN +L L++ D Y+L
 25 Sbjct: 70 TNDTIKNFKYSHISSDITFFRYFISDFIEQDKVYLDADIVVMGSLTELYQTDLISNYFL 129

Query: 122 AAVKDPHNE---EGGFNAGMLLANLELWREDGLTKTLKTAELERHVVKTDQDSILNI 177
 AAVKD + FNAGMLL N + WRE +T+ L +R+ + DQSILN+
 30 Sbjct: 130 AAVKDIISEKIYVNNHIFNAGMLLINNKWREHNTQFCFLSEKIYNSLFDADQSILNL 189

Query: 178 VCHNRWLSLNKINWF--QTYDVVSKYNHRSVLYLINIENRTPNIIFLTSDKPKWNENSVAR 235
 + ++W+ IN+ +N+ T + +Y YL ++ P IIR+ T KPW R
 35 Sbjct: 190 IFKDKWKLKLRGNVYLLGTLYLFFKYGKTRYLE-DLGETIPLIIHYVTEAKPWLINIFNTR 248

Query: 236 FRELAWYFYQLDFCQLTAK 254
 FR ++W+Y++L++ + K
 40 Sbjct: 249 FRNIYWFYELAWQDIYAK 267

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 40 vaccines or diagnostics.

Example 2027

A DNA sequence (GBSx2138) was identified in *Sagalactiae* <SEQ ID 6273> which encodes the amino acid sequence <SEQ ID 6274>. Analysis of this protein sequence reveals the following:

45 Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 50 bacterial cytoplasm --- Certainty=0.0417 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-2286-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2028

A DNA sequence (GBSx2139) was identified in *S. agalactiae* <SEQ ID 6275> which encodes the amino acid sequence <SEQ ID 6276>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -2.60      Transmembrane      306 - 322 ( 306 - 322)

----- Final Results -----
      bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF28363 GB:AF224467 putative glycosyl transferase [Haemophilus
      ducreyi]
      Identities = 88/259 (33%), Positives = 156/259 (59%), Gaps = 11/259 (4%)

Query: 7 VVLGGDYSYIRQIETTLKSLCVYHNLNLSIFPNQDIPQEWFLAMKDRVGQTNQIQDKL 66
      +VLA + SY I TT+K3+ +++++ ++ N+D P EMF + +++ + ++I D+K+
Sbjct: 10 IVLAANGSYSEYILTITKSLVLANKHIFVYLLNRDYPTEWFDILANKLAKLNSIIDIKV 69

Query: 67 FHDHLSPKWENKILRHINY-MTARYFIPQYISADTVLYLSDSLVVTINLDMFLPQISLN 125
      +D + K +HI+ T+ RYFI +I D V+YLD+D+VW +L L+Q + N
Sbjct: 70 TNDTIK--NFKTYSHISSDTFFRYFISDFIRQDKVIYLDADIVVMSUTELYQTDLSN 126

Query: 126 AYLAARV-----ALFGLGYGFRAGVNVINNRWRGEMNTIKLEKQKEIRMANESDQTI 180
      +LAARV ++ + FRAG+++IRN++NR+ N+T + ++K I + + DQ+I
Sbjct: 127 YFLAARVDIIESEKIYVNRHIFNAGMLLNNKQREHNTIQCLSLSEKIYNSLPDQDSI 186

Query: 181 LNRKFNQCVIYLDOTYFQIGFD-MGAADIGHKFIQDIPITPLEKIIHYISGIKPWQTLG 239
      LN +F+++ + L+ YN+ IG D + +++ D+ T +P IHHY + KFW +
Sbjct: 187 LNLIPDKKWLKLRGYNVLIGTDYLPFKYQKTRYLEDLGET-IPLIHHYNTKAPWLNIF 245

Query: 240 NMRLEVRVWHYNNLENSGI 258
      N R R ++W Y L W I
Sbjct: 246 NTRFRNTYWPYYELANWQDI 264
```

No corresponding DNA sequence was identified in *S. pyogenes*.

SEQ ID 6276 (GBS395) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 5; MW 47.4kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 83 (lane 8; MW 72kDa) and in Figure 177 (lane 5; MW 72kDa).

GBS395-GST was purified as shown in Figure 217, lane 7.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2029

A DNA sequence (GBSx2140) was identified in *S. agalactiae* <SEQ ID 6277> which encodes the amino acid sequence <SEQ ID 6278>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
```


-2287-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1633 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2030

A DNA sequence (GBSx2141) was identified in *S.agalactiae* <SEQ ID 6279> which encodes the amino acid sequence <SEQ ID 6280>. Analysis of this protein sequence reveals the following:

Possible site: 44

>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.16 Transmembrane 36 - 52 (36 - 52)

----- Final Results -----

bacterial membrane --- Certainty=0.1065 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10243> which encodes amino acid sequence <SEQ ID 10244> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP: AAC77330 GB: AE000508 orf, hypothetical protein [Escherichia coli K12]
 Identities = 75/260 (28%), Positives = 123/260 (46%), Gaps = 22/260 (8%)

Query: 6 VGLVLEGGGMRGLYTAGVLDAPLDAGIK-IDGIVSVSAGALPGVNFVSRQRERALRYNKK 64
 + LV EGGG RG++TAGVLD F+ A D + SAGA F+ Q A + +
 Sbjct: 25 IALVCEGGGQRGIFTAGVLDPEFNRAQPNPFDLVLTGSAGAQNLAFICNQPGYARKVIMR 84

Query: 65 YLSHPKMYSLRSWFRITGNFVNKD?---TYIEVPMKLD---VFDDEAFKSSIDFYVVA 116
 Y + ++ + R GN ++ D+ T ++P++D +PD S FY+ A
 Sbjct: 85 YTTKREFFDPLRFVRRGNLIDLWLVEATASQMLQNDTAARLFD-----SGKSFYCA 138

Query: 117 TMTSGKPEYFKLDSVFPQNEILRASSALEPVSKM-VWQKQKYLGGGLSDSIPVDFARG 175
 P YF + + +++RASSA+P + V +G YLDGG+SD+IPV A
 Sbjct: 139 CRQDYAPNYP-LPTQKNLWLVIRASSAIPGFYRSGVSLRGINLYLGGIDAIIPVKEAR 197

Query: 176 LGPDKLIVVMTRPLNYQKKESGR----LYKTLRYKYPNFVKTPASNYQQYNNLSLEKM 230
 G L+V+ T P P + L + + N V+ Y+ +EK
 Sbjct: 198 QGAKTLAVIRTVPQSMYTFPWFKMERMLGDSSLPVLNVLQHETSYRDIQOFIEKPP 257

Query: 231 SLEKTDLFAIRPSKSLVIG 250
 + +++ +P S+ +G
 Sbjct: 258 GKLRIPRIYPPKPLHSIALG 277

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8959> and protein <SEQ ID 8960> were also identified. Analysis of this protein sequence reveals the following:

-2288-

```

Lipid: Possible site: -1      Crend: 10
McG: Discrim Score:      -5.16
GVH: Signal Score (-7.5): -2.17
      Possible site: 44
>>> Seems to have no N-terminal signal sequence
ALOM program      count: 1 value:      -0.16 threshold: 0.0
INTERNAL Likelihood = 4.14      Transmembrane 36 - 52 ( 36 - 52)
PERIPHERAL Likelihood = 4.14      18
modified ALOM score: 0.53

```

-2289-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2032

A DNA sequence (GBSx2143) was identified in *S.agalactiae* <SEQ ID 6283> which encodes the amino acid sequence <SEQ ID 6284>. This protein is predicted to be transporter protein. Analysis of this protein

10 sequence reveals the following:

Possible site: 49

>>> Seems to have a cleavable N-term signal seq.

15	INTEGRAL	Likelihood = -6.85	Transmembrane	373 - 389 (370 - 395)
	INTEGRAL	Likelihood = -6.74	Transmembrane	168 - 184 (162 - 187)
	INTEGRAL	Likelihood = -6.32	Transmembrane	259 - 275 (257 - 280)
	INTEGRAL	Likelihood = -4.78	Transmembrane	286 - 302 (285 - 306)
20	INTEGRAL	Likelihood = -3.19	Transmembrane	55 - 71 (54 - 71)
	INTEGRAL	Likelihood = -2.97	Transmembrane	84 - 100 (79 - 101)
	INTEGRAL	Likelihood = -2.87	Transmembrane	311 - 327 (310 - 328)
	INTEGRAL	Likelihood = -1.44	Transmembrane	355 - 371 (355 - 371)
	INTEGRAL	Likelihood = -0.64	Transmembrane	108 - 124 (108 - 125)

----- Final Results -----

25 bacterial membrane --- Certainty=0.3739(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

30 >GP: AAC22759 GB: U32790 transporter protein (Haemophilus influenzae Rd)
 Identities = 139/391 (35%), Positives = 221/391 (55%), Gaps = 4/391 (1%)

Query: 6 INKNNRRLAIAIVASGTTDLNIMFLAFSMSTIITDLHLASAAQAGMIGTITNLMGLVQGL 65
 +N W+ALI + V G D +++ L F +S I DL+L+ AQ G + T T +G + GG+

35 Sbjct: 5 VNSYGNKALIGSAVGYGMDGFDLLILGFMLSAISADLNLTFAQGGSLVYWTILGAVFGGI 64

Query: 66 IPGLLADRYNKFVKFWITLIFSIINTGLVFPTTNSLYNIMRFIAGIGVGEYGLAIAM 125
 +PG L D+Y + +V WTLI+F++ TGL L I R IAGIG+GGE+GI +A+

40 Sbjct: 65 LPGALSDKYGRVRLVLTWILLFAVFTGLCAIAQGYWDLITYRTIAGIGLGEFGIGALALA 124

Query: 126 AGIVPTNKGRISSINIGAQVGSISSALLAGMLAPALGMRLFLPGLLPVLVLMQFA 185
 A P + +S + QVG + +ALL L P +GMRG+FL G+ P + +++

45 Sbjct: 125 ARANDARHRAKAASYVALGQVGVLAALLTPILLPHIGRMGLVGIPLAFVAFNPLRSH 184

Query: 186 VDDKIDILQYNTDADDEPLDI ---SIKALFDTFVLATQSLALMVMITVQIAGYFGKQW 241
 + + +I Q T + S + L + SL ++V+T+VQ GY+G+M W

50 Sbjct: 185 LHEPEITFQKQALSTQSGSPDTKLRSQLLIKDKATSKISLGIYVLTEVQNFQYQIMW 244

Query: 242 LPTIITQINLVSVNKSNNMIAITILGCMGLMFLVQGLDLKFGPRLVGCFMSSAICVYL 301
 LP + L S+ S I N T+ GM G+ +PQL D+ G + + F L + I + +

55 Sbjct: 245 LPNPLSKQLGFSLTSGIMTAVTVCGMMAGIWIQFQADIRGRKPSFLPLQGLGAVISIVV 304

Query: 302 FQFATTPMSMIIIGGAVGVFPVNGMFGAGYGMITRLYPHIRSTANNILNVGRAGIGPSS 361
 + T M++ GA +G FVNGM GYGA++ YP R+TA N++ N+GRA+GGF

Sbjct: 305 YSQLTDPDIMLLAGAFGLMFVNGMLGGYGALMAEAYPTBARATAQNVLNIGRAVGGPFGP 364

Query: 362 VIIGHILDVSNVSMVFLASLYIVSFLSML 392
 V+IG ++ + + Lh +Y++ L+ +

Sbjct: 365 VVVGSVVLAYSQTAIALLAITYVIDMLATI 395

-2290-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2377> which encodes the amino acid sequence <SEQ ID 2378>. Analysis of this protein sequence reveals the following:

```

Possible site: 39
5  >>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -8.92    Transmembrane    168 - 184 ( 162 - 188)
    INTEGRAL    Likelihood = -5.41    Transmembrane    286 - 302 ( 285 - 306)
    INTEGRAL    Likelihood = -5.15    Transmembrane    372 - 398 ( 362 - 394)
10  INTEGRAL    Likelihood = -3.45    Transmembrane    259 - 275 ( 257 - 276)
    INTEGRAL    Likelihood = -2.87    Transmembrane    311 - 327 ( 306 - 328)
    INTEGRAL    Likelihood = -2.81    Transmembrane    55 - 71 ( 51 - 71)
    INTEGRAL    Likelihood = -0.48    Transmembrane    108 - 124 ( 108 - 125)
    INTEGRAL    Likelihood = -0.37    Transmembrane    84 - 100 ( 84 - 100)

15  ----- Final Results -----
        bacterial membrane --- Certainty=0.4567 (Affirmative) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

20  An alignment of the GAS and GBS proteins is shown below.
        Identities = 306/402 (76%), Positives = 354/402 (87%)

Query: 1  MSPFINNNKNNRRLAIAIVASGTDLDNINFLAPSMSTIITDLHLSAAQAGWIGTITNLGM 60
MS L+++ N RAL+AAI ASGTDDLN+MFLAPSM+S+I+TDL LS Q GWI TITNLGM
25  Sbjct: 1  MSTLSLDYTNKRALVAIAASGTDDLNNVFLAPSMSSIMTDLGLSGTQQGWATITNLGM 60

Query: 61  LVGGLIFGLLADRYNKKFKVFKWITLIFSIATGLVFFTTNLSYLYIMRPIAGIGVGGEYGI 120
LVGGL+FGLLADR++KFKVFKWITL+FS+ATGL++FT +L YLY+MRPIAGIGVGGEY+
30  Sbjct: 61  LVGGLIFGLLADRHKKFKVFKWITLFLFSVATGLTYFTQSLFYLYIMRPIAGIGVGGEYGV 120

Query: 121  AIAINAGIVPTNKKGRISLINGIAQVQSISALLAGMLAPALGWRGLFLGGLFPIVLVL 180
AIAINAGIVP KNKR+SSLINGIAQ+GSISALLAGMLAP+LGWRGLFLGGLFPI+LV+
35  Sbjct: 121  AIAINAGIVPPSKNGRMSSLINGIAQVQSISALLAGMLAPSLGWRGLFLGGLFPIVLVI 180

Query: 181  WQQFAVDKIDLDQYNTDADDEPLDISIKALFDTFVLATQSLAMVMVTVOIAGYFGMN 240
WM A+DD+ I D Y + ++ I I L F T L Q+LALMVMVTVOIAGYFGMN
40  Sbjct: 181  WNTLAIDDKIWDHYGQSEECSPQIKINELFKTKSLTAQTLALMVMVTVOIAGYFGMN 240

Query: 241  WLPTIIQTNLNVSVKNSLWMIAITILGMCLGNLVFGQLLDKFGKPLVYGCFLSSAICVY 300
WLPTIIQT+LN+SVK+SSLWM+ATI+GMCLGNL FGQLLD FGKPL+Y FLL+S+ICVY
45  Sbjct: 241  WLPTIIQTSLNLSVKSSLWMVATIVGMCLGNLVFGQLLDGFGKPLIYSLFLASSAICVY 300

Query: 301  LQFQATTPMSMIIGGAVUGFFVNGMFPAGYGAMITRLYPHRIESTANNILNVGRAIGGFS 360
LQFQA +M SM+IGGA+VGFFVNGMFPAGYGAMITRLYPHRIESTANN+ILNVGRA+GGFS
50  Sbjct: 301  LQFQANSMSMVIIGGAVUGFFVNGMFPAGYGAMITRLYPHRIESTANNVILNVGRALGGFS 360

Query: 361  SVIIGMILNVSVKMWLFLASLYIVSFLSMISIKLKKRQY 402
SV IG ILD S +SMVM+FLASLY++SF +M SI QLK ++Y
Sbjct: 361  SVAIGSLDASGISKVMFLASLYIVSFGAMWSIGQKARKY 402

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2033

A DNA sequence (GBSx2144) was identified in *S.agalactiae* <SEQ ID 6285> which encodes the amino acid sequence <SEQ ID 6286>. This protein is predicted to be leucyl-tRNA synthetase (leuS). Analysis of this protein sequence reveals the following:

```

Possible site: 52
>>> Seems to have no N-terminal signal sequence

```

-2291-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3481 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5

A related GBS nucleic acid sequence <SEQ ID 10241> which encodes amino acid sequence <SEQ ID 10242> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAC00259 GB:AF008220 leucine tRNA synthetase [Bacillus subtilis]
 Identities = 569/835 (68%), Positives = 666/835 (79%), Gaps = 42/835 (5%)

15 Query: 10 YNKEIEBPKWQAFWADNHTFKTGTDAKPKFYALDMFFYPYSGAGLEVGHPEGYTADTILS 69
 + HKEIE KWQ +N +N TF T + K KFYALDMFFYPYSGAGLEVGHPEGYTADTILS
 Sbjct: 3 FQHKIEIKWQTYNLENTFTATLNNHEKQKPYALDMFFYPYSGAGLEVGHPEGYTADTILS 62

20 Query: 130 RFRAGQGRIVLHPMOMDAFGLPAEQYAMDTGNDPAETARNIANFKRQINALGFSYDMDR 129
 R KR QG++VLHPMOMDAFGLPAEQYA+DTGNDPA FT +NI NF+RQI ALGFSYDMDR
 Sbjct: 63 RMRKRQGYDVLHPMOMDAFGLPAEQYALDTGNDPAVFTKQNDINFRQIQALGFSYDMDR 122

25 Query: 130 EVNTIDPNYIKWTQWIFTKLYKGLAYEARVPMVVEELGTALANSEVLPGDTSERGYP 189
 E+NTIDP YKWTQWIF KLYKGLAY EVPMVN LGT +ANSEV+ DG SERGG+P
 Sbjct: 123 EINTIDPSYIKWTQWIFLKYKGLAYDEVPMVNMCFALGTVLANSEV+DGKSERGGP 181

30 Query: 190 VVRKPMRQWMLKITAYARILLEDLEEDWPFSEIKMQRRWIGKSGANVTFKVDTRDF 249
 V R+PM+QWMLKITAYA+RILLEDLEEDWPFSEIKMQRRWIG+S GA+V F + D F
 Sbjct: 182 VERRPMKQWMLKITAYARILLEDLEEDWPFSEIKMQRRWIGSGAHVHFDLDF 241

35 Query: 250 TVFTTRFDTLPGATYAVLAPHEALVDATTDAQAEAVYKQKQSLKSLDARTLAKET 309
 TVFTTRFDTLPGATY VLAPHEALV+ ITTA+Q EAV Y ++ KSLD RTDLAK KT
 Sbjct: 242 TVFTTRFDTLPGATYTVLAPHEALVENITTAQAEAVEYIKIQSLKSLDARTLAKET 301

40 Query: 310 GVWTGAYAINFVNGKSIPIWIADYVLA SYGTGAIMAVEAHDERRDSFAKQPNLDIIPVLE 369
 GV+TGAYAINFVNG+++P+WIADYVLA SYGTGA+MAVP HDERR+EFAK F L + V++
 Sbjct: 302 GVFTGAYAINFVNGKSLPIWIADYVLA SYGTGA+MAVPGHDERDPEFAKFTGLPVKHEVVK 361

45 Query: 370 GGNVSEAAFTEDGLHINSDFLDGLDKAAAIAMVENVLEABGVNEKVTYRLRDWLFPSRQR 429
 GGNVSEAA+T DG H+NSDFL+GL K AI K++ WLE G +KVTYRLRDWLFPSRQR
 Sbjct: 362 GGNVSEAAFTDGSHVNSDFLGLGLKQAELEKVIAMLESTKNGSEKVTYRLRDWLFPSRQR 421

50 Query: 430 YWGEFPIPIHWDGCTAVPSELEFLVLPVTKDIRPSGTGSEPLANLTDWLEVT-REDGV 488
 YWGEFPI+IHWEDGCTAVPE ELFL+LP T +I+PSGTGSEPLAN+ ++WVT E G
 Sbjct: 422 YWGEFPIVHWDGCTAVPSEELFLILLPKTDEIKPSGTGSEPLANKEWVEVDPEGTG 481

55 Query: 489 KGRRENTMPCWAGSWSYIYRIDPHNTEKLADEELLKQMLPVDIYVGGABRAVHLHLYA 548
 KGRRENTMPCWAGS WY+LRYIDPHN ++LA E L+LWLPVDI+Y+GGABRAVHLHLYA
 Sbjct: 482 KGRRENTMPCWAGS WYFLRYIDPHNFDQLASPEKLEKMLPVDIYVGGABRAVHLHLYA 541

60 Query: 549 RFWRKVLVYDLGVVPTKEPFQKLFNQCMILGTSYRDSRGALVATDKVEKRDGSPFHVETGE 608
 RFWRK LVD+GVVPTKEPFQKLF+NQCMILG E E
 Sbjct: 542 RFWRKVLVYDLGVVPTKEPFQKLFNQCMILG -----ENNE 575

65 Query: 609 ELBQAPAKMSKSLNVVNFDDVVEQYGAADTLRVYEMPMGFLDASIAWSSEGLSGSRFLD 668
 RMSKS NVVNFDD+V +GADTL+YEMPMGFLDASIAWS GL+G+R+FLD
 Sbjct: 576 -----RMSKSGNVVNFDDVVEQYGAADTLRVYEMPMGFLDASIAWSSEGLSGDARRFLD 628

70 Query: 669 RVYRLI-----TYKEITEHNSGALDKVYNEIVKAVTEQVDQMKNYATLAQWMPVNGAN 722
 RV+RL +I E L++VY+ETV VT+ ++FNT I+QWV+V+N A
 Sbjct: 629 RVNRLFIEDSGELNGKIVGAGETLRSVYHETVMKVTIHYGLRPFNVTIGSLQVFLNRAV 688

75 Query: 723 KSKDKLSDYANGVQQLIAPFAPHGLSEELWQVLTSAGQSISYVWPSYDSSKLVEINEITV 782
 K +L +Y +GFV+L+P APH: EELW+ L SG +I+Y WP YDE+KLV+R+E+EV
 Sbjct: 689 KATFLPKYEMKGFVLLSPVAPHLASEELWKLHSGS-TIAYEAWPVYDETKLVDEVEITV 747

80 Query: 783 VQIKGKVKAKLVAKDLRSKEIQDLALANRQVQARIAGDKIIVKLVKLVNIV 837
 VQ+ GRVKAKL V D ++E+L+ LA A+EKV+ ++ GK I R+IAPV KLVNIV

-2292-

Sbjct: 748 VQLNGKVKAKLQVPADATKQLRQLAQADEKVKKQLGKTKIRKIIIVPGKLVNIV 802

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6287> which encodes the amino acid sequence <SEQ ID 6288>. Analysis of this protein sequence reveals the following:

5 Possible site: 46
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.4358 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 815/833 (97%), Positives = 827/833 (98%)

15 Query: 7 NTFFNHEKEIEPKWQAFWADNHTFKGTGDASKPKFYALDMFFYPGAGLVGHGPEGYTATD 66
 MTFY+H IEPKWQAFWADNHTFKGTGDASKPKFYALDMFFYPGAGLVGHGPEGYTATD
 Sbjct: 1 NTDFYHTAIIEPKWQAFWADNHTFKGTGDASKPKFYALDMFFYPGAGLVGHGPEGYTATD 60

20 Query: 67 ILSRPFKRAQGHNLHFMGWDAPGLPABQYAMDGTGNDPAEPTAENIANFKRQINALGFSYD 126
 ILSRPFKRAQGH+LHFMGWDAPGLPABQYAMDGTGNDPAEPTAENIANFKRQINALGFSYD
 Sbjct: 61 ILSRPFKRAQGHNLHFMGWDAPGLPABQYAMDGTGNDPAEPTAENIANFKRQINALGFSYD 120

25 Query: 127 WDREVNITDIPNYYKWTQNIPTKLYEKGLAYEAEVPMVVEELGTAIANEVLPDGTSERG 186
 WDREVNITDIPNYYKWTQNIPTKLYEKGLAYEAEVPMVVEELGTAIANEVLPDGTSERG
 Sbjct: 121 WDREVNITDIPNYYKWTQNIPTKLYEKGLAYEAEVPMVVEELGTAIANEVLPDGTSERG 180

30 Query: 187 GYFVVRKPMRQMLKITAYAERLLEDLEEDVWPESI+KDWQRNWIKGSTGANVTFKVKDIT 246
 GYFVVRKPMRQMLKITAYAERLLEDLEEDVWPESI+KDWQRNWIKGSTGANVTFKVKDIT
 Sbjct: 181 GYFVVRKPMRQMLKITAYAERLLEDLEEDVWPESI+KDWQRNWIKGSTGANVTFKVKDIT 240

35 Query: 247 KDPTVFTTRPDTLFGATYAVLAPHAVALDAITTAQAEAAVAEYKRQASLKSLARITDLAK 306
 KDPTVFTTRPDTLFGATYAVLAPHAVALDAITTAQAEAAVA+YKRQASLKSLARITDLAK
 Sbjct: 241 KDPTVFTTRPDTLFGATYAVLAPHAVALDAITTAQAEAAVAEYKRQASLKSLARITDLAK 300

40 Query: 307 EKTGVWIGAYAINPVGKEIFVWIADYVLA SYGTGAINAVFAHDERDWEFAKQENLDIIP 366
 EKTGVWIGAYAINPVG E+FWIADYVLA SYGTGAINAVFAHDERDWEFAKQ LDIIP
 Sbjct: 301 EKTGVWIGAYAINPVGMEFWIADYVLA SYGTGAINAVFAHDERDWEFAKQFKDITIP 360

45 Query: 367 VLEGGNVEEAAFTEDGLHINSDFLDGLDKAAAIADVMLEBAEGVGNKVTYRLRDWLFS 426
 VLEGGNVEEAAFTEDGLHINS FLDGLDKA+AIADVMLEBAEGVGNKVTYRLRDWLFS
 Sbjct: 361 VLEGGNVEEAAFTEDGLHINSDFLDGLDKAIAIDVMLEBAEGVGNKVTYRLRDWLFS 420

50 Query: 427 RQRVWGRPIPIIHWDGTSVAVSESELPVLVLPVTKDIRPSTGSEPLAN+TDWLEVTRED 486
 RQRVWGRPIPIIHWDGTSVAVSESELPVLVLPVTKDIRPSTGSEPLAN+TDWLEVTRED
 Sbjct: 423 RQRVWGRPIPIIHWDGTSVAVSESELPVLVLPVTKDIRPSTGSEPLAN+TDWLEVTRED 480

55 Query: 487 GVKRRREINTMPQAGSSWYLYRIDPHNTEKLADRELLKQMLPVDIYVGGARHVALILL 546
 GVKRRREINTMPQAGSSWYLYRIDPHNTEKLADRELLKQMLPVDIYVGGARHVALILL
 Sbjct: 481 GVKRRREINTMPQAGSSWYLYRIDPHNTEKLADRELLKQMLPVDIYVGGARHVALILL 540

60 Query: 547 YARFWHKVLYDLGVVPTKEPFQKLFNQMIIGTSYRDSRGALVATDKVERKRGSGFFHVET 606
 YARFWHKVLYDLGVVPTKEPFQKLFNQMIIGTSYRDSRGALVATDKVERKRGSGFFHVET
 Sbjct: 543 YARFWHKVLYDLGVVPTKEPFQKLFNQMIIGTSYRDSRGALVATDKVERKRGSGFFHVET 600

65 Query: 607 GEELBQAFANKSKLKNVNFDDVVEQYGADITRVYEMFMGPELDASIAWSEEGLESGSKF 666
 GEELBQAFANKSKLKNVNFDDVVEQYGADITRVYEMFMGPELDASIAWSEEGLESGSKF
 Sbjct: 601 GEELBQAFANKSKLKNVNFDDVVEQYGADITRVYEMFMGPELDASIAWSEEGLESGSKF 660

Query: 667 LDRVYRLITTKREITEENSGALDKVYNETYKAVTEQVDQMKFNTALQALMVFNAAKBEK 726
 LDRVYRLITTKREITEENSGALDKVYNETYKAVTEQVDQMKFNTALQALMVFNAAKBEK
 Sbjct: 661 LDRVYRLITTKREITEENSGALDKVYNETYKAVTEQVDQMKFNTALQALMVFNAAKBEK 720

Query: 727 LFSDYANGFVQLIAPPFAMHIGENIAQVLTASQSGISYVFWPSYDKSKLVNRIIVVQIK 786
 LFSDYANGFVQLIAPPFAMHIGELIMQ LTASG+GISYVFWPSYDESKLVN++SIVVQIK

-2293-

Sbjct: 721 LFSDYAKGFVQLIAPFAPHLGRELWQALTASGESISYVPWPSYDESKLVENDVHIVVQIK 780

Query: 787 GKVKAKLVVAKDLRSRELQDLALANEKVQAEIAGKDIIVKLVAVPNKLVNIVVK 839
GKVKAKLVVAKDLRSRELQ+ALANEKVQAEIAGKDIIVKLVAVPNKLVNIV+K

5 Sbjct: 781 GKVKAKLVVAKDLRSRELQDLALANEKVQAEIAGKDIIVKLVAVPNKLVNIVIK 833

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2034

- 10 A DNA sequence (GBSx2145) was identified in *S. agalactiae* <SEQ ID 6289> which encodes the amino acid sequence <SEQ ID 6290>. This protein is predicted to be KIAA1074 protein. Analysis of this protein sequence reveals the following:

Possible site: 35
>>> Seems to have an uncleavable N-term signal seq
15 ----- Final Results -----
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20 A related GBS nucleic acid sequence <SEQ ID 8961> which encodes amino acid sequence <SEQ ID 8962> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 7
SRCFLG: 0
25 McS: Length of UR: 19
Peak Value of UR: 2.86
Net Charge of CR: 4
McS: Discrim Score: 10.27
GVH: Signal Score (-7.5): -3.61
30 Possible site: 31
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program count: 0 value: 2.12 threshold: 0.0
PERIPHERAL Likelihood = 2.12 7
35 modified ALOM score: -0.92

*** Reasoning Step: 3

40 ----- Final Results -----
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 45 No corresponding DNA sequence was identified in *S. pyogenes*.

SEQ ID 8962 (GBS117) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 8; MW 22.5kDa).

GBS117-His was purified as shown in Figure 200, lane 7.

- 50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2035

A DNA sequence (GBSx2146) was identified in *S. agalactiae* <SEQ ID 6291> which encodes the amino acid sequence <SEQ ID 6292>. This protein is predicted to be YirC (resE). Analysis of this protein sequence reveals the following:

- 5 Possible site: 28
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -10.88 Transmembrane 177 - 193 (173 - 196)
 INTEGRAL Likelihood = -4.09 Transmembrane 10 - 26 (5 - 29)
- 10 ----- Final Results -----
 bacterial membrane --- Certainty=0.5352 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
- 15 The protein has homology with the following sequences in the GENPEPT database.
 >GP:CAB15292 GB:Z99120 similar to two-component sensor histidine
 kinase [YvgA] [Bacillus subtilis]
 Identities = 106/379 (28%), Positives = 193/379 (50%), Gaps = 33/379 (8%)
- 20 Query: 92 DNHKCSHDIIIRYLTKRLMQISKEKDMFVTIKKKYTYVMTKYSGILVDSGSKKVPKA 151
 +N + S + L+ + ++ K D KKK Y + D +G V IKK
 Sbjct: 86 ENEEASSDKDLSSLSSTFHKVYKLDKQ--EAKKKRY---SADVNGEKVFFVFKKGLSV 140
- 25 Query: 152 QBSQFVHVFNS-----DITTYQHLITKINHFLVILVLTITPMLFMKRTFTGTSIRESIQ 205
 Q +++++ D+ YL L ++ + V++L++IP +++ + +
 Sbjct: 141 NGQSANMLSYALDSYRDDLAYT--LFKQLLFITIAVILLSWIPAINLAKY----LSRPLV 194
- 30 Query: 206 SVQTTISLWKVQGNHQSQKEIVPSDFDPLLESQEMANRIYQAESQRFQNASHEL 265
 S + ++ + ++ + K + L +EM ++ Q +E++R QN SH+L
 Sbjct: 195 SFEKHVKRI--SEQWDDPVKVRKDEIGKLGHTIEENRQKLVQKDETERTLQLNISHDL 252
- 35 Query: 266 RTPMLSIQGYTEGVQBGII---DAELAHSVILQESKNNKQLVDDIILSKLD--SNLSQD 320
 +TP+M I+GYT+ ++GI D E VI E+ K+++ + D++ L+KLD + Q
 Sbjct: 253 KTPVMVIRGYTQSKDGIFFKDLNTVDVIRCEALKLEKKIKOLLYLTQKLVYAKQKQV 312
- 40 Query: 321 KDEPSLAEELLNSIIAYFKPLANKQKISITYRFPKHEKLLK-GNEBELQRAINNLSNALR 379
 D FS+ E+ +I K A K+ +++ D E +L G+ E + + NIL N+R
 Sbjct: 313 KDMPSIVEVTEIEVERLK-WARKE---LSWEIDVEEDILMGDPEQNNKLENILENQIR 368
- 45 Query: 380 YAVSHIEISYT---NQKLTISNDGPAISKEDLPYIPDRFYKHGGCTGIGLAWTCIEIK 435
 YA + IEIS N +TI NDGP I E L ++ F KG G+ GIGL++ K I+
 Sbjct: 369 YAEKIEISMKQDDRNIVITIKNDGPHIEDMLSSLYEPNKGKKGEGPGLSIVKIRILT 428
- 45 Query: 436 QHHNIIARSDSTSTFTFI 454
 H +I E+D T ++ I
 Sbjct: 429 LHKQSIISIENDKTGVSYRI 447

There is also homology to SEQ ID 1178.

- 50 SEQ ID 6292 (GBS279) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 7; MW 54.5kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 6; MW 79.4kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2295-

Example 2036

A DNA sequence (GBSx2147) was identified in *S.agalactiae* <SEQ ID 6293> which encodes the amino acid sequence <SEQ ID 6294>. This protein is predicted to be two-component response regulator (intraA). Analysis of this protein sequence reveals the following:

```

5   Possible site: 37
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10  bacterial cytoplasm --- Certainty=0.1706 (Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10239> which encodes amino acid sequence <SEQ ID 10240> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB05663 GB:AP001513 two-component response regulator [Bacillus halodurans]
Identities = 87/220 (39%), Positives = 124/220 (55%), Gaps = 4/220 (1%)

20 Query: 11 IYFADDEKNIRDLVVFLEHDGFTVRAFETGDLLEAYKNQKFDLVILDIMMPGTSLDW 70
   I DDE ++R+LV +L +GF V ETGD ++ + + DLV+LD+NM +G
   Sbjct: 7 ILIVDELDLRLVLYSLRKGFAVYTAETGDEALKRLEQEPMDLVLDVNMNDMDGPTA 66

Query: 71 MKSIRQYDNIPILMLTARSDVDFTAFNLQTDYFTKPFSPKILSLHVKALPKRLDEKA 130
   K IR + IPIIMLTAR + D + +G DDY KPFSP +L ++ +R
25 Sbjct: 67 CKEIRAFSQIPIIMLTARGSEDDKVMGLQIGADDYIVKPFSPRELVARIEVALRRTQGIQ 126

Query: 131 IKNDTYQYQFLDLTLDTKRIALLSNEEMFLTKTEFDLLVLIKPEAFSRETLNRIWG 190
   +DT Y+F +L + R ++ +E+ LTK E+D L+ L+E F+RE L +R+WG
30 Sbjct: 127 QVDDTYGRFNEIRIQPSGRKVPVMQGISLTKKEYDLVFLLEHGRVFTREHLMDRLMG 186

Query: 191 FDDIES--RAVDDTIKRLKKPKQYHSQVSIKTVWGYGFK 228
   D + R VD IK LR KK + IKTVMG G+K
35 Sbjct: 187 MTCQGLTGLRTVTHIKTLRLKLP--ADRFIKTVWGVGYK 224

```

35 There is also homology to SEQ ID 3260.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2037

40 A DNA sequence (GBSx2148) was identified in *S.agalactiae* <SEQ ID 6295> which encodes the amino acid sequence <SEQ ID 6296>. Analysis of this protein sequence reveals the following:

```

Possible site: 55
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL Likelihood = -2.18 Transmembrane 1568 -1584 (1568 -1585)
45  INTEGRAL Likelihood = -0.16 Transmembrane 338 - 354 ( 338 - 354)

   ----- Final Results -----
   bacterial membrane --- Certainty=0.1871 (Affirmative) < succ>
   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
50  bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10237> which encodes amino acid sequence <SEQ ID 10238> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-2296-

>GP:AAG09771 GB:AF243528 cell envelope proteinase [Streptococcus thermophilus]
Identities = 797/1594 (50%), Positives = 1056/1594 (66%), Gaps = 39/1594 (2%)

- 5 Query: 21 MNTKQRFIRKYKLGAVSLVGLITFFLOGITNVAADSVINKPSDIAVEQVKQDSPTS-IA 79
M K + FS+RYK+G VSVLLG + F +VAAD + + + VR V D+ S A
Sbjct: 1 MKKKESTFSIRKYKLGITVSVLLGAVFLFAGAPVAADSLTSLV-ETKVRATVPDAIVBSSEA 59
- 10 Query: 80 NETPTNN--TSSALASTAQDNLVTKANNSTPTQFVARSHSQATSTFSFVANOQPVSTQEE 137
+E+P +++ +T+ D T ++ + S + P S ++
Sbjct: 60 SESPVVEELVITSVRATSTDTTITDNKEKTPGSKALRNSAN:KVEITTPQAVETPAISSEK 119
- 15 Query: 138 VSKETPLTKQLNAVSKTSPAIKSKET--PQNDISNKIITVPRVMTGYKKEGSTVVAIIDSGLD 195
V + K +A+ +T +E PQNDISN IITVPRVW +GYKKEGSTVVAIIDSGLD
Sbjct: 120 VESEE--KLSVADETTATITQREAKPQNDISNTIITVPRVWYSGYKKEGSTVVAIIDSGLD 177
- 20 Query: 196 INHDALQLNDSTKAYQNSQOMNAKAKAGINYGKWNKVI FGHNYVDVNTLEKVRST 255
++RD L ++D + AKY++E+++ AAK AGI YG+H+N+KV+FG+NYVDVNT LKE
Sbjct: 178 VDRDVLHISLSTAKYKSEKRIEAAKEAAGITYGWFNDKVPVGVNYVDVNTVLKEKDR 237
- 25 Query: 256 SHGMVTSIATNPSSKDTNELIYGVAPQAQVMKRVFSDEKRGTPALYKAIEDAVKL 315
SHGEVTSIAT NP++ +L+YGVAPQAQVMKRVFSD K TG ALYKAIEDAVKL
Sbjct: 238 SHGMVTSIATNPITQPVAGQLMYGVAPEAQVMKRVFSDLKATGAALYKAIEDAVKL 297
- 30 Query: 316 GADSNILSGGANGSLVADDRLIKALEMARLAGVSVVIAAGNDGTFGSGASKPSALYDP 375
GADSNILSGGANG+VN ++ + A+E AR AGVSVVIAAGNDGTFGSG S PSA YPD
Sbjct: 298 GADSNILSGGANGSVVNMNSHVTAIEAARRAGVSVVIAAGNDGTFGSGHNSPSADYDP 357
- 35 Query: 376 YGLVGPSPSTAREAISVASYNNTITLVNKVFNI IGLNENRNLNGLAAYADPKVSDKTFEFG 435
YGLVG+PSTA +AISVASYNNTI+ +KV NIIGLENN +LN G+++ +P+ S E+G
Sbjct: 358 YGLVGAPSTAHDAISVASYNNTITVGSKVINI IGLNENADLNYGKSSFDNPEKSPVPTFEIG 417
- 40 Query: 436 KQYDVYFVGKQNDNDYKDKTLNGKIALIERGDIITFKKVVNAINHAGVGAIIIPNKAQEA 495
K+Y+YV+ G G +D+ L GK+ALI+G ITH+K+ NA GAVG +IFN++
Sbjct: 418 KEYETVYTAGIQASDFGLDITGLKALIKKGTITFSEKIANATAAGAVGVVITFNSRGEA 477
- 45 Query: 496 NLTMSLDPEASAI PAITFQKEGVDLAKNRYKIVFNINKNQANPAAGVLSDFPSGWLTA 555
X+M LD A AIP++F EFG+ LA R+YKI FNN + + NP AG+LSDFPSGWLTA
Sbjct: 478 NVSMQLDDTALAI PSVFIPLFEGEALAAANSYKIAFNKTDIRPNPEAGLLSDFPSGWLTA 537
- 50 Query: 556 DGQLKPLDAPQGSIIYAINDNEYDMAGTSMASPHVAGATALVQKYLKHEPLKKGDI 615
DG+LKPDL+AR3G+IYAAINDN+Y M GTSMASPHVAGA LVQKYL +P +I
Sbjct: 538 DGQLKPLDAPQGSIIYAINDNEYDMAGTSMASPHVAGAVLVQYLLATYPTKSPQEI 597
- 55 Query: 616 ERTVKYLLMSTAKAHLNKITGAYTSPRQQAGIIDVAAAVOTGLYLIGBNYNGSVTLGN 675
E VK+LLMSTAKAH+NK+T AYTSRQQAGIILD AAA+ TGLILT GE+ YGS+TLGN
Sbjct: 598 EALVKKLLMSTAKAHVKNKTTAYTSRQQAGIITDAAALSTGLIIT-GEQDGSITLGN 656
- 60 Query: 676 IKDKISFDVTVBNINKVARDLHYTTTLATDQKDGFTVLAQQLGTFPKKTRIRIEPQQTQ 735
++D SF VT+HNI K L+Y+L TD + L + + + + + + +
Sbjct: 657 VEDTFSFTVLHNTINEDKTLNISTVLTTTAKRILDLHGTSISRDGMRKVTKANBST 716
- 65 Query: 736 TITIDVSKTHMLKKVMNNGYFLBGTVRPDPVDGGEVLSIPYVFGKGEFQNLVLEK 795
+T+I++D S + + L +M NGI+LEG+VRFD D G+++SIPYVFG+GEFQNL VLE+
Sbjct: 717 TVTINVDASSFAELGTGLMNGSYLYLGGFVRFPDVAQDGIIVSIPYVFGKGEFQNLVLEK 776
- 70 Query: 796 SIYKLVANKKQKYPQ--KQINEVPGSHVYALMTTSEPIYSTGTSTPIQLKALGSYK 853
IY L+A+ + GFYE+P Q N V S YT L+T S+E IYSTD S +K L+K+K
Sbjct: 777 PTYNLLADGKGKYPFPEVPTAQNTVDISHHYTLVIGSGTELYSTDKRSDSALITLGTFF 836
- 75 Query: 854 SIDGKWLLODKQOPHILASINDDQNDQAVKGVFLRNFNLRKAVKPSQVDFNKL 913
+ G ++L+LD+ G+PHILASIN D NQD++ KQVFLRN+ +L A VY ADI KFL
Sbjct: 837 NKNGTFLVLELDESGKPHILASINDDQNDQSLVFKGVFLRNTYDLVASVYADDTERTNPL 896
- 80 Query: 914 NVASPAQGDQKRYSGNFKPNKSTFLYDTNKGITITDGIPLBKGKRYKYTVYSDVPKSP 973
W S PQ+GDQN YSGN +NPKS+ +Y TEW GT +DG L DGK+YVITY S VRG+
Sbjct: 897 WESQPSQSGKSHYSGNFKPNKSSIIYPTENWNTSDGNALADGKCYVITYSSKVPGRV 956
- 85 Query: 974 QQNVFDITLDRQAPLITATYDKDRRI FKARPAVEHGESGIFREQVPLYKKDKGHYNSV 1033

-2297-

Q M+PD+ +DR++P +TTATYD+ F RPA+E GESG++REQVFFYL D G ++
 Sbjct: 957 QTMIPDVIIDRESFVITATYDETNTFNPRAIEKSGSLYREQVFILVADASS-VTTI 1015

5 Query: 1034 LRQGGEGILVEDNKVFIPKQKDGSPILFKEVNDPSHVYTTVEDYAGRLVSAKLEDLINI 1093
 + V DKIVF+ Q DGSF LP ++ D S YTYVEYAGH+ K+E+LI+I
 Sbjct: 1016 PSLIKNGGVTVSINKV+VQNDGDSFTLPLDLADISKFYTTVEDYAGNISYEKVENLISI 1075

Query: 1094 GNGKGLVNVKVFSPFLNSMVDIDPSYSVKDDKGNIIKK-QHHGKDLNLLKLPFGTYTFDL 1152
 GN+ GLV V + + NS V I PSYSV D+ G I+ + + D ++LKLFPFGTYTFDL
 10 Sbjct: 1076 GNEKGLVNTLDDKDTNSFVPLTSPYSVIDETGKIVARLPYAGDTSVLKLPFGTYTFDL 1135

Query: 1153 FLYDEERANLISPKSVTVTISEKDSKLKDLVLFKVNLLKKAALLVEFDKLLPKGNATVQLVTK 1212
 FLYD E ++L VTI E +S +V F V L KA LL++ D LLP G+T+QLVT
 15 Sbjct: 1136 FLYDTWSSLAGETKAVVITILEENSTARVNFYVTLAKDKANLLIDIALPSSGSTIQLVTA 1195

Query: 1213 TMTVVDLPEKATYSPTDYGNLFPVGDYRLNVLPSPGYSTLENLDDLNSVKEDQVNLTKLT 1272
 + LP A YS TDYK +PVG Y + TLP GY LE LD V+V +Q N+ KLT
 Sbjct: 1196 DQQAQLPNAKYSKTDYKGFVPGTYTTLPTLPBGYEFLEELD---VAVLANQSNVKKLT 1252

20 Query: 1273 LINKAPLINALABOTDIITQPFYNAGTHLKNNYLANLEKACTLIQNKVEQTSIDNAIAA 1332
 LINK L +AE + +YNA L+ Y LE A + N+ Q +D+A+A-
 Sbjct: 1253 LINKVALKELIABLAGLESTARYNAPSELQATAYAKALEDANAVYANKHNAQAQVDSALAS 1312

Query: 1333 LRESQALNGKETUTSLIAKATIAETIKGNVQFVNASPLSQSTYINQVLAKNLLQKN 1392
 L +R+ LING+ TD L + T + N+ + NA Q Y V+ +L+ N+
 25 Sbjct: 1313 LVAAREQLNGQATDEKELIAEVSNTYTPQANFYTYNAENTQIAIDYTVASQAQLVLNQEN 1372

Query: 1393 VTQSEVDKALENLDIAKQLNGHETDYSGLHHMIIKANVLKQTSSKYQNASQFAKENYNN 1452
 VTQ+ V++AL +L AK L+G +TD S L + ++VLK T +KY NAS+ K+ Y+
 30 Sbjct: 1373 VTQAVVNALADLLAAKANLDGQKTDISALRSASVSVSLKATDAKYLNASENQATDQ 1432

Query: 1453 LIKKAELLNRRHTQAEVELLNQIKATEQELDG----RDRVSSAANSYSQSLNDNDSLN 1508
 ++ A+ +L + A+QA V++ L + + + ELDG + N + D ++
 35 Sbjct: 1433 AVEAAKAILVDSEASQASVQLANLTSQAQALDGVATSDNAKEPANTATDKKDEGYTV 1492

Query: 1509 TTPIN-----PP-----NQQAIFPKGMTKESEVAKGVLVGTSQTDNKKVETLK 1555
 PI+ PP N I +K + + + L + + NQ+ + +L
 Sbjct: 1493 PPPIDSEIVDQAPPVKDTGNSHVPQIGQK-PNPQPTLPRPVTLQASLSNPKQKQVTQL 1551

40 Query: 1556 PKTGSTPKITYTILLFSLNGLATIKLKSIRK 1569
 P TGE+ K L ++GL T+ L SI+R
 Sbjct: 1552 PNTGENDTK----YILVPQVITGLQTL-LVSIIR 1560

A related GBS gene <SEQ ID 8963> and protein <SEQ ID 8964> were also identified. Analysis of this protein sequence reveals the following:

45 Lipop: Possible site: -1 Crend: 10
 SRCFLG: 0
 MG: Length of UR: 1
 Peak Value of UR: 2.55
 Net Charge of CR: 4

50 MG: Discrim Score: 2.60
 GVH: Signal Score (-7.5): -0.78
 Possible site: 35
 >>> Seems to have a cleavable N-term signal seq.

55 Amino Acid Composition: calculated from 36
 ALOM program count: 1 value: -0.16 threshold: 0.0
 INTEGRAL Likelihood = -0.16 Transmembrane 318 - 334 (318 - 334)
 PERIPHERAL Likelihood = 2.54 1161
 modified ALOM score: 0.53
 60 icml HYPID: 7 CFP: 0.106
 *** Reasoning Step: 3
 ----- Final Results -----
 65 bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2298-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 1535-1539

5 The protein has homology with the following sequences in the databases:

50.5/67.5% over 1583aa

Streptococcus

s thermophilus

GP|9963932| cell envelope proteinase Insert characterized

ORF01603 (361 - 5070 of 5370)

GP|9963932|gb|AAG09771.1|AF243528_1|AF243528(1 - 1584 of 1585) cell envelope proteinase {*Streptococcus thermophilus*}

 $\%Match = 41.2$

%Identity = 50.4 %Similarity = 67.4

Matches = 794 Mismatches = 498 Conservative Sub.s = 267

255 285 315 345 375 405 435 465
KRALGTVLNLQPNRL**KFRKL*KILIPYVLIVFVIIMLQEKSEIPMTKQPSIRKYKLGAVSVLLGTLPFLGGITVAA
| : | :: ||| : | |||| : | |||| : | : | : |
MKKKETFSLRKYKIGTYSVLLGAVFLFAGAPSVAR
10 20 30

DSVINKPSDIAVEQQVKDSFTS-IANETPT--NNTSSALTAQENLVTKANNSPITQFVAESHQSQTETFSPPVAMQPV
:
:
DE-LTSLVETKVETAPDAIVSESGASFPWVELDVTSVETS-DVTITDNEEETPGSEALEANSANTEVETTPQAVETAPA

50 60 70 80 90 100 110

[illegible][illegible]

1200 1230 1260 1290 1320 1350 1380 1410
 RVFSDEKRTGTPALYYKAIEDAVKLGADSNINSLGGANGSLVNADDRLIKALEMARLAGVSVVIAAGNDGTFGSGASKPS
 RVFSDLQATTGAALYYKAIEDAVKLGADSNINSLGGANGSVNNNENVTAAIEARRAGVSVVIAAGNDGTFGSGHNSPS
 290 300 310 320 330 340 350

[illegible]

1680	1710	1740	1770	1800	1830	1860	1890
YKDKTNGKIALIGERDITPTKKVVNAINHGAVGAIIPNNKAGEANLTMSLDPRASATPAIPTQKSGFVWIAKNKYKIVF							
: : : : : : : : :	:	:	: : : : : : : : :	: : : :	: : : :	: : : :	: : :
FDGLDTGKLALIKRGTITPSEKIANATAAGVGVVIPNSRPGBANVMQLDDTALIASPVFIPLFFGEALANSYKIAFL							
450	460	470	480	490	500	510	

1920 1950 1980 2010 2040 2070 2100 2130
NNIKNKANPNPAGVLSDFSSWGLTADQQLKFDLSAPGSGSIYAAINEYDMMISGTSMASPHVAGATLWKQYLKEHPEL
| : : | : : | : : | : : | : : | : : | : : | : :
NNETIRFPNEAGLLSDFSSWGLSADGEKFDLIAAPGAAYAINENYDMMGSTSMASPHVAGAVALWKQYLILATPYTK

530 540 550 560 570 580 590

2160 2190 2220 2250 2280 2310 2340 2370
KKGDIERTVRYLLMSTAKAHLNKGDTGAYTSPROOGAGIIDVAAAVOTGLYLTTGGENNYGSVTLGNIKDKISFDVTVHTN